LECTURE NOTES IN GEOINFORMATION AND CARTOGRAPHY

LNG&C

Georg Gartner · William Cartwright
Michael P. Peterson (Eds.)

Location Based Services and TeleCartography

Springer

Lecture Notes in Geoinformation and Cartography

Series Editors: William Cartwright, Georg Gartner, Liqiu Meng, Michael P. Peterson

Georg Gartner • William Cartwright •
Michael P. Peterson

Location Based Services and TeleCartography

With 214 Figures

Springer

Editors:

Prof. Dr. Georg Gartner
Cartography and Geo-Mediatechniques
Department of Geoinformation and
Cartography - TU Vienna
Erzherzog-Johannplatz 1
1040 Vienna, Austria
Email: georg.gartner@tuwien.ac.at

Prof. Dr. Michael P. Peterson
Department of Geography/Geology
University of Nebraska at Omaha
Omaha, NE 68182 USA
Email: mpeterson@unomaha.edu

Prof. Dr. William Cartwright
School of Mathematical and Geospatial
Sciences
RMIT University
Melbourne
Australia
Email: william.cartwright@rmit.edu.au

ISBN 10 3-540-36727-6 **Springer Berlin Heidelberg New York**
ISBN 13 978-3-540-36727-7 **Springer Berlin Heidelberg New York**

Library of Congress Control Number: 2006937680

Springer is a part of Springer Science+Business Media
springeronline.com

Cover design: deblik, Berlin
Production: A. Oelschläger
Typesetting: Camera-ready by the Editors
Printed on acid-free paper 30/2132/AO 54321

Preface

This book provides for the first time a general overview of research activities related to location and map-based services. These activities have emerged over the last years, especially around issues of positioning, spatial modelling, cartographic communication as well as in the fields of ubiquitious cartography, geo-pervasive services, user-centered modelling and geo-wiki activities. The innovative and contemporary character of these topics has lead to a great variety of interdisciplinary contributions, from academia to business, from computer science to geodesy. Topics cover an enormous range with heterogenous relationships to the main book issues. Whilst contemporary cartography aims at looking at new and efficient ways for communicating spatial information the development and availability of technologies like mobile networking, mobile devices or short-range sensors lead to interesting new possibilities for achieving this aim. By trying to make use of available technologies, cartography and a variety of related disciplines look specifically at user-centered and context-aware system development, as well as new forms of supporting wayfinding and navigation systems. Contributions are provided in five main sections and they cover all of these aspects and give a picture of the new and expanding field of Location Based Services and TeleCartography.

Georg Gartner, Vienna, Austria
William Cartwright, Melbourne, Australia
Michael Peterson, Omaha, USA

Table of Contents

Georg Gartner
LBS and TeleCartography: About the book 1
1 A series of Symposiums on LBS and TeleCartography 1
2 Progression of Research 3
2.1 Terms 3
2.2 Elements 4
3 Structure of the book 6
3.1 The section "General Aspects" 6
3.2 The section "Positioning" 7
3.3 The section "Modelling and Awareness" 8
3.4 The section "Visualization and Cartographic Communication" ... 9
3.5 The section "Applications" 10
4 Summary 11

Section I : General Aspects

Jonathan Raper
Design constraints on operational LBS 13
1 Introduction 14
2 Two LBS that work 15
2.1 Pedestrian navigation: KDDI EZ Naviwalk 16
2.2 Mobile multimedia location guide: Camineo guides 17
3 What these working LBS show us we need in future 19
3.1 Concepts of context and activity 19
3.2 Context architectures 20
3.3 Positional fusion and rationalisation 21
3.4 Application development 22
4 Conclusions 23

Bin Jiang, Xiaobai Yao
Location Based Services and GIS in Perspective 27
1 LBS: definitions, characteristics, and application prospects 27
2 Modeling for LBS 30
2.1 User needs and modeling 30
2.2 Location modeling 31
2.3 Context modeling and adapting 33
2.4 Geospatial data processing and modeling 34
3 Research challenges for LBS 35
3.1 Naïve users and next-generation GIS 37
3.2 Spatio-temporal analysis and mining of mobile geospatial data 38
3.3 On-the-fly generalization and visualization 39
3.4 Interoperability issues 40
3.5 Privacy and social issues 41
4 Conclusion 42

Markus Uhlirz
A Market and User View on LBS 47
1 Location Based Services 47
1.1 Mobile Positioning and Location Based Services 48
1.2 Summary of Positioning Methods and Accuracies 49
2 Market aspects 50
2.1 Market trends 50
2.2 Value Chain for LBS 52
2.3 Regulatory and Legal Aspects 53
2.4 Ownership of Location Data & Privacy Issues 54
3 The User's View 54
3.1 Cartographic User interface 54
3.2 Suitability to the Use Case 55
3.3 Critical Success Factors for LBS from User's perspective 56
4 Conclusion 58

Gerhard Navratil, Eva Grum
What makes Location Based Services fail? 59
1 Introduction 59
2 Examples 60
2.1 LBS for emergency calls 60
2.2 LBS for dealer shows 61
3 Technical Solutions and their Pitfalls 62
3.1 Access to the Service 62
3.2 Processing the Inquiry 63
3.3 Transfer of Instructions 64

4 Legal Restrictions 64
4.1 Access to the Service 65
4.2 Processing the Inquiry 65
4.3 Transfer of Instructions 66
5 Usability 66
5.1 Access to the Service 66
5.2 Processing the Inquiry 67
5.3 Transfer of Instructions 68
6 Combined Influences 69
6.1 Technology – Usability 69
6.2 Technology – Legality 70
6.3 Legality – Usability 70
7 Discussion and Conclusions 70

Michael Peterson
The Transition from Internet to Mobile Mapping 73
1 Introduction 74
2 Contrasts in Development between Wired and the Mobile Internet .. 76
2.1 The Role of Government 76
2.2 Content Development 78
2.3 Compatibility with Existing Media 79
2.4 Developing a user base and paying for information 80
2.5 Free maps 81
2.6 Competition from wired Internet map use 82
3 Changing Patterns of Map Use 82
3.1 Navigation 82
3.2 GPS integration 83
3.3 People Tracking 84
4 Mental maps and wayfinding 85
4.1 Mental map formation 85
4.2 Mental map formation in wayfinding 86
4.3 Creating a Permanent Dependence 86
5 Conclusion 87

Takashi Morita
Theory and development of research in ubiquitous mapping 89
1 Map Communication 90
1.1 Three elements 90
1.2 Origin of Map Communication 91
1.3 Scheme of Map Communication 91
2 Mapping World 92
2.1 From "Map" to "Mapping" 93

2.2 Egocentric mapping 94
3 Fundamentals of ubiquitous mapping 95
3.1 Changes in the information and communication environment ... 95
3.2 Ubiquitous nature of maps 96
3.3 Difference between GIS and Ubiquitous Mapping 97
4 Research in ubiquitous mapping 98
4.1 Elements of context awareness 98
4.2 Applications in Tokyo 103
5 Conclusion 105

Section II : Positioning

Edward Verbree, Sisi Zlatanova
Positioning LBS to the third dimension 107
1 Introduction 108
2 The need for 3D LBS anytime, everywhere 108
3 Limitations in 3D positioning technologies 111
4 Opportunities of combined GPS-Galileo positioning for urban environments 113
5 Discussion 116

Günther Retscher
Altitude Determination of a Pedestrian in a Multi-storey Building .. 119
1 Introduction 119
2 Overview of indoor location systems 120
3 Altitude determination of a pedestrian using a barometric pressure sensor 123
4 Sensor tests: locating the user on the correct floor in a multi-storey building 123
5 Determinations of the sensor drift 124
6 Determination of a characteristic curve for the barometric pressure sensor 125
7 Determination of the height in the building 126
8 Concluding remarks and outlook 127

Rudolf Pailer et.al.
Terminal-Centric Location Services in IP Multimedia Subsystem ... 131
1 Introduction 132
2 The IP Multimedia Subsystem - IMS 133

3 Location Mechanisms and Standards 135
4 IMS Presence Location Client Architecture 136
5 IMS Presence Location Performance 139
6 Related Work 140
7 Conclusion and Future Work 141

Michael Thienelt et.al.
WiKaF - A Knowledge-based Kalman-Filter for Pedestrian Positioning 143
1 Introduction 143
2 System architecture 145
2.1 Sensors 145
2.2 Knowledge-based filtering 146
2.3 Central KALMAN-filter for optimal multi-sensor integration ... 148
3 Further workflow of WiKaF 150
3.1 Test areas 150
3.2 Workflow of WiKaF 150
4 Conclusions and perspectives 151

Andreas Eichhorn
Map-independent positioning of land vehicles with causative modified motion equations 153
1 Motivation 154
2 Multi-sensor system 155
3 KALMAN-filter algorithms with causative motion equations 157
3.1 Fundamentals 157
3.2 Causative modified motion equations 157
4 Creation of a 'Position Finding Module' 159
5 Results from different test-drives 161
6 Conclusions 163

Section III : Modelling and Awareness

Steffen Volz
Shortest Path Search in Multi-Representation Street Databases 165
1 Introduction 166
2 Related Work 167
3 A Short Introduction to Nexus 168
4 The Concept of MRep Relations 170

4.1 MRep Relations 170
4.2 Generating MRep Relations 171
5 An Approach for Shortest Path Search in MRep Databases 172
5.1 Generating an integrated navigation graph 172
5.2 Determining Possible Paths 173
5.3 Calculating the Costs 174
5.4 Realization of the Approach 175
6 Summary and Outlook 176

Inessa Seifert et.al.
Region-Based Representation for Assistance with Spatio-Temporal Planning in Unfamiliar Environments 179
1 Motivation 179
2 Collaborative Assistance with Spatio-Temporal Planning 181
2.1 Need for assistance 181
2.2 Partially unformalized constraint systems 182
2.3 Dealing with the complexity of geographic information 183
3 Hierarchies in Human Mind 183
3.1 Structural aspects of spatial environments 184
3.2 Human spatial problem solving strategies 184
4 Hierarchies in Artificial Spatial Systems 185
5 Regionalize and Conquer 186
5.1 Region-based assistance with spatio-temporal planning 188
6 Outlook and Future Work 189

Mohammad Malek et.al.
A Logic-Based Foundation for Spatial Relationships in Mobile GIS Environment 193
1 Introduction 194
2 Related Work 195
3 Preliminaries 196
4 Algebraic and Topological Structure 197
5 Expressivity Power 199
5.1 Qualitative Geometry 200
5.2 Coaching 200
6 Conclusion and Further Work 201

Kai-Florian Richter
From Turn-By-Turn Directions to Overview Information on the Way to Take 205
1 Introduction 205
2 Context-Specific Route Directions 207

3 Overview Information on the Way to Take 209
4 Generating Overview Information: An Outline 210
5 Summary 214

Masatoshi Arikawa, Kouzou Noaki
Geocoding Japanese Walking Directions using Sidewalk Network Databases 217
1 Introduction 217
2 Sidewalk Network Databases 218
3 Natural and Formal Route Descriptions 221
3.1. Structure of Walking Directions 221
3.2 Simple Model of Walking Directions 221
3.3 Grammar of Formal Route Statement (FRS) 222
6 Conclusion 229

Tumasch Reichenbacher
The concept of relevance in mobile maps 231
1 Introduction 231
2 The concept of relevance in other disciplines 233
3 Relevance in mobile cartography 234
3.1 Relevance and Context 234
3.2 Relevance types 235
4 Relevance assessment and relevance measures 236
4.1 Utility functions 236
4.2 Information retrieval functions 237
4.3 Fuzzy sets 237
4.4 Observation-based approaches 237
4.5 Geographic information relevance assessment 238
5 Applications of relevance in mobile cartography 239
6 Relevance visualisations 242
7 Conclusions 244

L. Tiina Sarjakoski et.al.
A Knowledge-Based Map Adaptation Approach for Mobile Map Services 247
1 Introduction 248
1.1 Background 248
1.2 Previous research 249
2 The GiMoDig- map service 250
3 Map Specification Knowledge Base (MSKB) 253
3.1 Context and context parameters 253
3.2 Map specifications 256

3.3 Implemented algorithm 257
3.4 Meeting the requirements of context modelling 258
4 Map Specification Tool 259
5 Concluding remarks 262

Albrecht Weiser, Alexander Zipf
A visual editor for OGC SLD files for automating the configuration of WMS and mobile map applications 265
1 The OGC SLD Specification - a formal representation of maps 266
2 The ArcMap2SLD Converter 267
3 Evaluating the SLD support of WMS servers 270
4 SLDs for configuring mobile maps 274
5 Summary and Outlook 276

Rainer Simon et.al.
Towards Orientation-Aware Location Based Mobile Services 279
1 Introduction 279
2 Point-to-Discover 280
3 Technological Requirements 283
3.1 Selection Quality 283
3.2 3D Environment Models 284
4 Orientation-Aware Mobile Devices 285
4.1 Magnetic Compass 285
4.2 Tilt Sensor 285
4.3 Related Work 286
5 Creative Histories Mobile Viewer 287

Patrick Luley et.al.
Geo-Services and Computer Vision for Object Awareness in Mobile System Applications 291
1 Introduction 292
2 User Scenario 293
3 System architecture 294
4 Visual Object Recognition for Object Awareness 297
5 Geo Services 298
6 Conclusions 299

Section IV : Visualisation and Cartographic Communication

Verena Radoczky
How to design a pedestrian navigation system for indoor and outdoor environments ... 301
1 Introduction ... 301
2 Applicability of various multimedia presentation forms ... 303
2.1 Maps ... 303
2.2 Floor plans ... 304
2.3 Verbal guidance ... 305
2.4 Images ... 306
2.5 Videos ... 307
2.6 3D presentation ... 307
2.7 Online services ... 308
3 Integration of active and/or passive landmarks ... 308
4 Additional Design Goals ... 310
5 Conclusions ... 313

Ken Francis, Peter Williams
Dancing_without_gravity: A story of interface design ... 317
1 Introduction ... 318
2 Positioning_our_work ... 318
3 Dancing_without_gravity ... 319
4 Our_approach ... 320
5 The_work ... 321
5.1 Art_direction ... 322
5.2 Map_design ... 322
7 Interface/information collapse ... 325
8 Conclusion ... 328

William Cartwright
Landmarks and the perception of a space in 3D-worlds ... 329
1 Introduction ... 330
2 Overview ... 330
3 Evaluation ... 334
4 Results ... 336
4.1 Level of detail ... 336
4.2 Landmarks ... 341
4.3 Inclusions ... 341
5 Relevance of research results to LBS and Small-screen devices ... 343
6 Conclusion ... 343

Michael Wood et.al.
Comparing the effects of different 3D representations on human wayfinding 345
1 Introduction 346
2 Some preliminary studies 348
3 The current project: Background and experimental strategy 349
3.1 The stimuli used in the experiments 350
3.2 Experiment 1: Pilot Study 351
3.3 Procedure 351
3.4 Results 352
3.5 Discussion 353
4 Conclusion 354

Jürgen Döllner et.al.
An Approach towards Semantics-Based Navigation in 3D City Models on Mobile Devices 357
1 Motivation 357
2 Related Work and Challenges of Mobile 3D City Models 358
2.1 Mobile 3D Rendering 359
2.2 Mobile 3D Interaction 359
2.3 Standardization and Distribution 361
2.4 Digital Rights Management 361
3 Sketch-Based Navigation Commands 361
3.1 Real-Time Interaction vs. Selective Interaction 361
3.2 Concepts of Sketch-Based Navigation 362
3.3 Sketch-Based Navigation Vocabulary 363
4 Client-Server Architecture for 3D Visualization 365
4.1 Server System 365
4.2 Client System 367
5 Conclusions 367

Dieter Schmalstieg, Gerhard Reitmayr
The World as a User Interface: AR for Ubiquitous Computing 369
1 Introduction 369
2 Mobile augmented reality platform 372
2.1 Hardware and software 372
2.2 User Interface 373
2.3 Tracking infrastructure 375
2.4 Applications 377
3 First experiences with large scale data models for AR 380
4 Automating visualization generation 381
4.1 Example: Gas utility company 382

4.2 Example: Pedestrian navigation 383
4.3 Adaptive visualization engine 383
4.4 Semantic reasoning engine 385
4.5 Building visualizations 386
4.6 System integration 387
4.7 Example: Tracking target objects for navigation........ 388
5 Conclusions and future work 389

Tobias Höllerer et.al.
"Anywhere Augmentation": Towards Mobile Augmented Reality in Unprepared Environments 393
1 Introduction 394
2 From Mobile Augmented Reality to "Anywhere Augmentation" ... 395
3 Previous Work 398
3.1 Wearable Systems........ 398
3.2 Feature Extraction 400
4 System 402
4.1 Calibration 402
5 Annotation Interface 403
5.1 Corners 405
5.2 Edges 406
5.3 Regions 407
6 Feature Extraction........ 407
6.1 Corners 409
6.2 Edges 409
6.3 Regions 410
7 Discussion........ 412
8 Conclusion 413

Rex Cammack
Open Content Web Mapping Service: A Really Simple Syndication (RSS) Approach 417
1 Introduction 417
2 Portable Maps 418
3 Syndication 420
4 Syndication for TeleCartography 422
5 Internet Syndication........ 424
6 RSS and Atom 425
7 TeleCartography through RSS Feeds 427
7.1 Base Map 428
7.2 Thematic Data 429
8 Conclusions 430

Section V : Applications

Carlo Ratti et.al.
Mobile Landscapes: Graz in Real Time 433
1 Introduction 434
2 The data 436
3 The real-time mapping system.... 439
4 The results 440
4.1 Cellphone traffic intensity 440
4.2 Traffic migration 441
4.3 Traces of registered users 441
5 Discussion.... 442
6 Conclusions 443

Rein Ahas et.al.
The Spatial Accuracy of Mobile Positioning: Some experiences with Geographical Studies in Estonia.... 445
1 Introduction 446
2 Mobile positioning and data collection.... 446
2.2 Collecting positioning data.... 448
2.3 Positioning data 450
3 Data.... 451
4 Accuracy of mobile positioning in the Estonian GSM network.... 453
4.2 Calculated positioning bias.... 454
4.3 Comparison of mobile positioning and GPS data in 2004 455
4.4 Comparison of two methods for calculation of central point of positioning sectors.... 457
5 Discussion and conclusions 458

Natalia Andrienko, Gennady Andrienko
A Framework for Decision-Centred Visualisation in Civil Crisis Management.... 461
1 Problem statement 462
2 Actors and their information needs.... 463
3 Knowledge.... 466
4 Meta-information.... 468
5 Knowledge sources.... 471
6 Examples 472
7 Conclusion.... 472

Erich Wilmersdorf
Providing an Information Infrastructure for Map Based LBS - The approach of the City of Vienna 479
1 Introduction 480
2 General requirements for a LBS infrastructure 480
2.1 Data requirements of LBS 481
2.2 GIS functionality 481
2.3 TeleCartography 482
3 The contribution of the urban information system for LBS 482
3.1 Fundamental data bases for LBS 482
3.2 The digital Geo-data Warehouse 485
3.3 Change Management of Geodata 488
4 Rendering Geographic Services 489
4.1 Analytical features 489
4.2 Real Time Cartography 489
4.3 Providing Web Services 490
5 Conclusion 491

David Bennett et.al.
MoGeo: A location-based educational service 493
1 Introduction 494
2 Principles to guide the application of *MoGeo* in higher education.. 496
3 *MoGeo* Technology 497
4 Teaching the four traditions of geography with *MoGeo* 500
4.1 The spatial tradition 501
4.2 The area studies tradition 504
4.3 The human/environment tradition 505
4.4 The earth science tradition 506
5 Future work 507
6 Conclusions 508

Barend Köbben et.al.
Bata Positioning System - A real time tracking system for the world's largest relay race 511
1 The Batavieren relay race 512
2 BPS: The Bata Positioning System 512
3 Basic requirements and setup 513
4 BPS 1.0 – the protoype 514
5 BPS 2.0 – the current system 517
6 BPS 3.0 prototype – Development of a real time LBS 519
7 Conclusion 523

Karl Rehrl et.al.
Smartphone-based information and navigation aids for public transport travellers 525
1 Introduction 526
2 Requirements 528
2.1 Mobile trip planning and management 528
2.2 Detailed modelling of pedestrian networks 528
2.3 Navigation and guidance for public transport travellers 529
2.4 Continuous positioning 529
3 Related Work 530
4 Design of the personal travel companion 531
4.1 Route calculation and modelling of interchange facilities 532
4.2 Navigation in complex public transport interchange buildings 533
4.3 Positioning 536
5 Prototypical implementation 539
6 Conclusion 541

Lilian Pun-Cheng et.al.
EASYGO – A public transport query and guiding LBS 545
1 Introduction 545
2 An overview of EASYGO LBS 547
3 EASYGO Subsystems and route searching 548
3.1 The Public Mode Subsystem 549
3.2 The Address Subsystem 549
3.3 The Map and Chart Subsystem 549
3.4 Computation of Optimal Routes 549
4 Accuracy of mobile positioning 550
5 Accuracy Assessment Results 552
6 Concluding remarks 554

Siegfried Wiesenhofer et.al.
Mobile City Explorer: An innovative GPS and Camera Phone Based Travel Assistant for City Tourists 557
1 Introduction 558
2 Application scenarios 559
2.1 Service registration 560
2.2 Starting a personalized tour 560
2.3 Guiding the tourist and route adaptation 561
2.4 Collecting impressions 562
2.5 Recognizing sights by taking photos 562
2.6 Revising and sharing travel impressions 562
3 System architecture and components 563

3.1 Smart Guide 563
3.2 Multimedia Geo-Server 566
3.3 Object Recognition 568
3.4 Travel Diary 570
3.5 Integration Platform 572
4 Summary 572

Byoung-Jun Kang et.al.
Development of Cultural Inheritance Information System using LBS Technologies for Tourists 575
1 Introduction 575
2 Motivation and Site Information 576
3 System Architecture 577
4 System Implementation 578
4.2 User interface 578
4.3 Main service 1 579
4.4 Main service 2 580
4.5 Main service 3 580
5 Conclusion 581

Michaela Kinberger et.al.
LWD-Infosystem Tirol – visual information about the current avalanche situation via mobile devices 583
1 Introduction 584
2 Topic 584
3 Conception and implementation 586
4 Results 587
4.1 Maps for mobile devices 588
4.2 Maps representing the current air temperature and wind 589
4.3 Maps representing the regional avalanche danger scale 590
5 Outlook 590

László Zentai, Antal Guszlev
Spatial tracking in sport 593
1 Overview 593
2 Rally 594
3 Orienteering 595
4 Soccer 599
5 Ballooning 601
6 Sailing 603
7 Conclusion 604

LBS and TeleCartography: About the book

Georg Gartner

Department of Geoinformation and Cartography, Vienna University of Technology, Austria
georg.gartner@tuwien.ac.at

Abstract

This book is based on a series of symposiums on Location Based Services and Telecartography that have been held since 2002 at the Vienna University of Technology. The meetings themselves were a response to technological developments in miniaturizing devices for telecommunication, computing and display and an increased interest in both incorporating cartographic presentations on such mobile devices and developing services that are specific to a particular location. The broad variety of disciplines involved in this research and the differences in approaching the basic problems is probably typical of a developing field of interdisciplinary research. However, some main areas of research and development in the emerging area of LBS and Telecartography can now be identified. The contributions to this book are mainly selected from the papers of the 3rd Symposium on LBS and TeleCartography and reflect the main areas of interest: positioning, modelling and awareness, visualisation and cartographic communication and application development.

1 A series of Symposiums on LBS and TeleCartography

Research and business activities in the field of applying cartographic presentations on mobile devices (TeleCartography) and developing innovative services, where the location of a mobile device becomes a “variable” of an information system (LBS), have increased internationally since the mass-market availability of mobile devices, satellite positioning and telecommunication infrastructure. The Cartography Department of the Vienna University of Technology has been interested in these developments from the early stages, which have led to various research projects, student activities, multidisciplinary cooperations and the organization of a series of symposiums on LBS and TeleCartography in Vienna. The success of the first symposium in 2002 (Kelnhofer et al 2002) and the increasing activities in various disciplines related to LBS and TeleCartography was the major impetus for setting up a second symposium in 2004 (Gartner 2004) and a third symposium in 2005 (Gartner 2005), being meant as forum for bringing together experts from academia and business as well as representatives of different disciplines.

The symposiums on Location Based Services (LBS) and TeleCartography have been organized by the Institute of Cartography and Geo-Media Techniques (IKGeoM)[1] of the Vienna University of Technology (TU) in close cooperation with the Commission on Maps and the Internet of the International Cartographic Association (ICA) and with a collaborative interest from the ICA Commission on Ubiquitous Cartography, the Working Group 4.1.2 of the International Association of Geodesy and the Working Group V TC 2 of the International Society for Photogrammetry and Remote Sensing. As the world authoritative body on cartography, the International Cartographic Association attempts to advance the use of maps in society and science. The ICA Commission on Maps and the Internet, formed in 1999, brings together international specialists in the field of Internet mapping and disseminates information to a broader audience on new developments and major areas of research.

As the activities in the field of LBS and TeleCartography can be understood as an expansion of methods for Internet mapping and techniques for the mobile Internet, the ICA Commission on Maps and the Internet has a common interest in supporting the meetings in Vienna, Austria, dedicated to the issues of LBS and TeleCartography.

The 3rd Symposium on Location Based Services and TeleCartography held in Vienna from 28 - 30. November 2005 with the purpose of offering a forum for research-driven activities related to the context of location and map-based services. Such activities emerged in the last years especially around issues of positioning, spatial modelling, and cartographic communication as well as in the fields of ubiquitous cartography, geo-

[1] Research Group Cartography and Geo-Media techniques within the Department of Geoinformation and Cartography since 1.1.2004

pervasive services, user-centred modelling or geo-wiki activities. The innovative and contemporary character of the conference leads to a great variety of contributions in terms of interdisciplinarity. Presenters representing 18 countries with backgrounds varying from academia to business, from computer science to geodesy cover an enormous number of topics with a heterogeneous relation to the conference main topic.

While contemporary cartography examines new and efficient ways on communicating spatial information, the development and availability of technologies like mobile networking, mobile devices or short-range sensors lead to interesting new possibilities of achieving this aim. Cartographers and researchers from a variety of disciplines try to make use of the available technologies by looking specifically at user-centred and context-aware system developments as well as new forms of supporting wayfinding and navigation systems.

Selected results of the interesting contributions to the Third symposium on LBS and TeleCartography, mirroring the main elements and trends within this field, are brought together in this book.

2 Progression of Research

2.1 Terms

The key terms being used in this book reflect the iterative way of applying terms and names to developments in the context of new technologies. In general a flexible approach in terms of defining the frame of the area of interest has been used, thus the main terms are applicable to most of the contributions.

- Cartographic Model: Presentation model of geospatial data, derived from the geo data model by means of cartographic generalization and symbolization.
- Location Based Service: Services that exploit knowledge about where an information device user is located.
- TeleCartography: Distribution of cartographic presentation forms via wireless data transfer interfaces and mobile devices.
- Ubiquitous Cartography: Ability for users to create and use maps in any place and at any time to resolve geospatial problems.

In modern cartography, the main focus is on understanding the processes and methods of “how to communicate spatial information efficiently”. In this concern, the “responsibility” of cartography exceeds the creation of cartographic presentation forms, but is rather focused on understanding the relations within the "whole system" of communicating spatial information, including the user, the models and the transmission processes. The engagement of modern cartography in fields like LBS and TeleCartography and the various multidisciplinary approaches including cartography have to be seen in this context.

Telecommunication infrastructure (mobile network), positioning methods, mobile in- and output devices and multimedia cartographic information systems are prerequisites for developing applications– that incorporate the user's position as a variable of an information system. Normally, cartographic presentation forms are involved in integrating geospatial information into such a system.Thus, the resulting system can be called a *map-based location based service* (LBS). This chapter discusses the elements of a map-based LBS, outlines the main research topics and describes some experiences in the context of conceptual design and the development of map-based LBS.

2.2 Elements

A system can be called a Location Based Service (LBS), when the position of a mobile device – and therefore the position of the user – is somehow part of an information system. The derivable types of applications in this context can be stated as heterogenous and include simple and text-based applications, which use the *cell ID* (unique identification of the cell of a telecommunication network) for a rough positioning (e.g., Which petrol stations are there around me?) to map-based multimedia applications including routing functionality. In this context, different names for the context of telecommunication infrastructure, location-based applications and cartography are used. In addition to *mobile cartography*, *ubiquitous cartography* the author proposes the term *TeleCartography*, to be understood as issues involved by the distribution of cartographic presentation forms via wireless data transfer interfaces and mobile devices.

Independent from the level of complexity of the system architecture, every map-based LBS needs some basic elements to handle the main tasks of positioning, data modelling and information presentation.

- Positioning

The determination of the position of a mobile in/output device is a direct requirement for every system to be called LBS. Positioning has to be adequate to the service in terms of a dependent relationship and adapted to the tasks. For various applications the necessary level of accuracy needed can be served by the cell-ID of a telecommunication network and the thus derivable position, that gives an accuracy of positioning between 50 and 100 meters in urban areas. For navigation purposes – in particular in the context of pedestrian navigation – the accuracy demands increases to values of at least 25 meters and less. For indoor navigation, the requirements for the position determination are even greater.

Various methods of positioning are available for different levels of accuracy:

- satellite-based positioning,
- positioning by radio network,
- alternative methods,
- combinations.

- Modelling and Presentation of Information

The possibilities of transmitting and visualising geospatial information in context of a determined position is primarily restricted by the limitations of the particular mobile device. The basic conditions of the cartographic communication process have to be fulfilled when using map-based LBS: The cartographic model has to be clearly perceivable while it is permanently scale-dependent and has to present the task-dependent appropriate geometric and semantic information.

This fact in combination with restrictions in size and format of current mobile devices leads to different levels of solutions for presenting information within map-based LBS:

- Cartographic presentation forms without specific adaptions
- Cartographic presentation forms adapted to specific requirements of screen display
- New and adapted cartographic presentation forms
- Multimedia add-ons, replacements and alternative presentation forms

- Users and Adaptation

Experiences with LBS developments have lead to various suggestions for a more user-centered system conception. Modelling parameters in the context of the *user* and the *usage situation* are seen as fundaments of such user-adequate attempts, that can be summarized as *concepts of adaptation*.

The adaptation of cartographic visualisations in this context can be understood as, for example, the automatic selection of adequate scales, algorithms for adequate symbolization, or even the change to text-only output of information in case of shortcomings in size or resolution of the output device. Adapting to the needs of the user results from the user profile, selected in advance from a list or entered manually by the user to influence the graphical presentation (size of lettering, used colours) or to provide pre-defined map elements. Adapting the visualisation to the situation would include the current day and time and the speed of travel.

- Towards Ubiquitous Environments
The development of technologies like telecommunication infrastructures, wireless networks, radio frequency identification or innovative displays like electronic paper can all be seen as ways of developing a ubiquitous environment, where location-based services and ubiquitous cartographic applications can be applied.

3 Structure of the book

The book is structured in five main sections, reflecting the results of the symposiums and the main progression of research:

- General Aspects
- Positioning
- Modelling and Awareness
- Visualisation and Cartographic Communication
- Applications

3.1 The section "General Aspects"

The first section "**General Aspects**" concerns the various issues related to overall developments in LBS, TeleCartography and Ubiquitous Cartography.

In chapter 2 **Jonathan Raper** assesses the needs for the development of LBS to be successful. The major design constraints of operational LBS are identified, analysed and discussed. **Bin Jiang,** et.al., present an overview of LBS modeling regarding users, locations, contexts and data in chapter 3. The LBS modeling endeavors are crossexamined with a research agenda of geographic information science and some core research themes are briefly speculated. In chapter 4 **Markus Uhlirz** reviews the market position of location-based services in the typical mobile operator's service portfolio and discusses possible reasons leading to this situation. **Gerhard Navratil,** et.al., investigate possible pitfalls for Location-Based Services in chapter 5. The categories of pitfalls are identified as technical possibilities, legal restrictions, and usability. In chapter 6 **Michael Peterson** deals with the transition from Internet Mapping to mobile mapping and examines differences in the development of wired Internet and wireless Internet mapping usage and applications. In chapter 7 **Takashi Morita** defines theory and elements of ubiquitous cartography. In his understanding, ubiquitous mapping refers to the use and creation of maps by users anywhere and at any time. It is strongly influenced by advances in information technology, such as the development of wireless systems, high-density data storage and broadband communication, that can be seen as stimulating and facilitating dynamic and personalized mapping.

3.2 The section “Positioning”

In the second section “**Positioning**” various aspects of determining the position of a mobile user are investigated.

In chapter 8 **Edward Verbree and Sisi Zlatanova** focus on the possibilities and limitations in 3D positioning with respect to LBS. Challenges of a combined GPS/Galileo positioning in urban environments are addressed by a simulation study on the availability of sufficient signal reception within city areas. **Günther Retscher**’s chapter 9 briefly investigates methods to determine the altitude of a pedestrian in a multi-storey building. In this case it can be recommended to augment the position determination system with a barometric pressure sensor for direct observation of height differences. In chapter 10 **Rudolf Pailer,** et.al., state, that current access networks determine user location in a network-centric manner and present a terminal-centric location enabler that integrates seamless with the existing IP Multimedia Subsystem presence architecture and interoperates with network-centric location mechanisms. **Michael Thienelt,** et.al., present a new knowledge-based approach for the improvement of position quality with a Kalman-Filter approach for improving precise, reliable and preferably positioning of mobile users that is available everywhere in chapter 11. In chapter 12 **Andreas Eichhorn** presents a map-independent module for the autonomous positioning of vehicles within the framework of Location Based Services (LBS). The integration of the absolute position information (GPS) and the relative information (differential odometer and gyro) is realized with a central KALMAN-filter.

3.3 The section “Modelling and Awareness”

In the third section “**Modelling and Awareness**” aspects of modelling geo-data, determining routes and other elements for LBS and navigation

Systems and the consideration of user-dependent variables for developing adaptive or context-aware systems is discussed.

In chapter 13 **Steffen Volz** discusses problems concerning spatial data integration resulting from the fact that the same real world objects are stored in multiple, inconsistent representations (MRep) in different spatial databases. An approach to find an optimized solution for the navigation in MRep street networks by generating explicit relations between multiple representations and by exploiting them during shortest path search is described. **Inessa Seifert,** et.al., present in chapter 14 a cognitively motivated approach to an interactive assistance system for spatio-temporal planning tasks. Since mental spatial problem solving is known to be based on hierarchical representations, it is argued for region-based representation structures that allow for structuring a complex spatio-temporal problem such that exploring and communicating alternative solutions to a given problem are easily enabled. In this way, spatio-temporal constraints can interactively be dealt with to find appropriate solutions for a given planning problem. In chapter 15 **Reza Malek,** et.al., attempt to provide a paradigm to treat moving objects in mobile GIS environment. An idea based on space and time partitioning is suggested. A logic-based framework for representing and reasoning about qualitative spatial relations over moving objects in space and time is proposed. **Kai-Florian Richter** concentrates in chapter 16 on in-advance route directions that provide a coarse overview on the way to take. He discusses the required properties of these directions in order for them being useful and present some first methods to generate such instructions. In chapter 17 **Masatoshi Arikawa,** et.al., propose an advanced method of geocoding for walking directions in documents using daily local expressions. A core schema of sidewalk network databases is described and a formal route statement to represent and process walking directions by means of a computer is proposed. In chapter 18 **Tumasch Reichenbacher** looks on the concept of relevance and its potential for application to mobile cartography. A basic model for the determination of geographic relevance is postulated and the application modes of relevance to mobile cartography are elaborated. **Tiina Sarjakoski,** et.al., introduce in chapter 19 a Map Specification Knowledge Base approach in order to be able to utilise and manage various map specifications for mobile maps. A Map Specification Knowledge Base is a set of specification rules, properties and assignments constituting the system's knowledge about the map to be delivered to the mobile user. In chapter 20 **Albrecht Weiser,** et.al., describe the potential of the OGC Styled Layer Descriptor (SLD) specification for describing the design and contents of a map in a formal way through the use of a strict XML schema. Originally designed in order to allow clients to specify the look of maps served by WMS servers it is shown how it can also be used for configuring WMS as well as mobile map applications running locally on PDAs. In chapter 21 **Rainer Simon**, et.al., present an approach on how orientation sensors built into mobile devices can enable a new paradigm for mobile service discovery and use. In conjunction with 3D models of urban terrain, mobile devices can act as virtual pointers to services and information anchored at geographic locations such as buildings or landmarks. **Patrick Luley**, et.al., focus on contextual indexing for mobile system applications in chapter 22. Computer vision can be used to extract object information and thereby enables object based indexing which can finally be applied to extract a semantic description of the environment.

3.4 The section "Visualization and Cartographic Communication"

In the fourth section "**Visualization and Cartographic Communication**" various presentation forms are discussed in the context of LBS and TeleCartography. The range of presentation forms include maps, 3D-presentations as well as augmented reality presentations.

In chapter 23 **Verena Radoczky** discusses the use of active and passive landmarks, as well as multimedia presentation forms and general design goals for a combined indoor/outdoor pedestrian navigation service. **Ken Francis,** et.al., try to give a designer's perspective on map interfaces as a sense of control and understanding for users in chapter 24. In chapter 25 **William Cartwright** reports on a research programme that is building and testing a 3D tool for community collaborative decision-making. It concentrates on the stage of the research that focuses-on the question: How does a priori knowledge of an area change navigation and exploration abilities when 'moving through' and exploring the virtual world? **Michael Wood,** et.al., report on comparing the effectiveness of conventional topographic maps and computer-based geovisualisation systems as aids to navigation in wilderness mountain areas in Scotland in chapter 26. In chapter 27 **Jürgen Döllner**, et.al., outline a novel approach for user navigation in complex virtual 3D city models on mobile devices. Users navigate within the virtual 3D city model by sketching navigation commands in the perspective view on the mobile client. **Dieter Schmalstieg,** et.al., discuss the possibilities of augmented reality (AR) as a ubiquitous user interface to the real world in chapter 28. A mobile AR system can constantly provide guidance to its user through visual annotation of the physical environment. In chapter 29 **Tobias Höllerer,** et.al., present a novel approach to "Anywhere Augmentation" based on efficient human input for wearable computing and augmented reality (AR). **Rex Cammack** looks in chapter 30 at using RSS protocols to create an environment that causes more geospatial portal users to become long-term users. RSS is a syndication method for delivering content to users. By pushing content to users on a timely basis one can address the three-stated user loss factors.

3.5 The section "Applications"

In the final section "**Applications**" a broad variety of developed prototypes, attempts and services are described as well as the development of ideas of using LBS and TeleCartography services as data acquisition tools.

In chapter 31 **Carlo Ratti**, et.al., describe results of the "Mobile Landscapes: Graz in Real Time" project. Three types of maps of the urban area of Graz, Austria, were developed and seem to open the way to a new paradigm in urban planning: that of the real-time city. **Rein Ahas**, et.al., describe in chapter 32 some aspects of mobile positioning accuracy when using mobile positioning data in geographical studies. The method used in Estonia is named the Social Positioning Method and uses locations of mobile phones and the personal characteristics of phone owners for studying human behaviour. In chapter 33 **Natalia Andrienko**, et.al., describe a framework for intelligent visualisation to be applied in emergency situations for supporting the crisis management personnel and for informing and instructing the population. **Erich Wilmersdorf** describes in chapter 34 the components of a city infrastructure suitable for LBS applications, able to promote public but also private LBS projects by offering dynamic content and standardized Web-interfaces for geo-data extraction (WMS, WFS). In chapter 35 **David Bennett**, et.al., present a location-based service for geographic education, based on contextually-aware computing and the assemblage of technologies to integrate important elements of the classroom and computer laboratory into a field based learning environment. **Barend Köbben**, et.al., describe in chapter 36 the Bata Positioning System, a real time system for tracking moving objects. In chapter 37 **Karl Rehrl**, et.al., propose a personal travel companion on Smartphones for assisting multimodal travellers by focusing on the topics of mobile journey planning, pedestrian route calculation, navigation and guidance in public interchange buildings and seamless transition between indoor and outdoor positioning. **Esmond Mok**, et.al., describe in chapter 38 a public transport query and guiding LBS for the city of Hong Kong. In chapter 39 **Siegfried Wiesenhofer**, et.al., present a concept and architecture to provide location aware travel assistance for city tourists based on a combination of wide spread conventional mobile phones with built in cameras and GPS/GALILEO receivers. **Byoung-Jun Kang**, et.al., describe in chapter 40 a prototype of cultural inheritance information system that provides tourists with interactive and dynamic information using Location-Based Service. In chapter 41 **Michaela Kinberger**, et.al., describe a complex and database driven online decision support system for visualization and analysis of current avalanche relevant factors in the Tyrolean Alps. **Laszlo Zentai**, et.al., describe in chapter 42 the possibilities of Location-based Services in different kinds of sports.

4 Summary

The concept of the book has been outlined in this chapter. It is discussed that the results of three international symposiums at the Vienna University of Technology mirror the heterogenous variety of disciplines and backgrounds - which have contributed aspects and issues within the “umbrella terms” of Location-based Services and TeleCartography. The main areas of research, that can be derived from these results, are identified and the structure of the book is explained.

References

Gartner, G. (2004): Location Based Services and TeleCartography. Proceedings of the 2nd Symposium on LBS and TeleCartography. Geowissenschaftliche Mitteilungen, 66, Wien 2004.

Gartner, G. (2005): 3rd Symposium on Location Based Services and TeleCartography. Geowissenschaftliche Mitteilungen, 74, Wien 2005.

Kelnhofer, F. & M.Lechthaler, K. Brunner (2002): TeleKartographie und Location Based Services. Beiträge zum 1.Symposium TeleKartographie und LBS. Geowissenschaftliche Mitteilungen, 58, Wien 2002.

Section I General Aspects

Chapter 2	Jonathan Raper Design Constraints on operational LBS
Chapter 3	Bin Jiang, Xiaobai Yao Location Based Services and GIS in Perspective
Chapter 4	Markus Uhlirz A Market and User View on LBS
Chapter 5	Gerhard Navratil, Eva Grum What makes Location Based Services fail?
Chapter 6	Michael Peterson The Transition from Internet to Mobile Mapping
Chapter 7	Takashi Morita Theory and development of research in ubiquitous mapping

Design constraints on operational LBS

Jonathan Raper

Geographic Information Centre, Department of Information Science, City University, London, UK
raper@soi.city.ac.uk

Abstract

In the last 15 years of accelerated digital technology evolution a well-established model of product and service development known as the hype cycle has been seen to effectively characterise the progress of specific technologies from birth to acceptance. LBS need to be assessed against this template and currently it can be argued that they are lying in the 'Trough of disillusionment'. In order to ensure that LBS move through the cycle towards the 'Plateau of productivity', the emerging LBS inter-discipline needs an assessment of what is needed for LBS to succeed.

Drawing upon the experience of an EU research project 'Webpark' (http://www.webparkservices.info/) and two UK research projects Locus (http://www.locus.org.uk/) and LBS4all (http://www.lbs4all.org/), this paper will examine a design template for LBS focussed on operational requirements and constraints. This template focuses on the following inter-locking requirements for:

- Mobile device memory, communication interfaces and positioning support
- operating systems for mobile devices and their customisation potential
- development platforms supporting real-time positioning and network-agnostic behaviour
- graphics technologies with performance and integration with development platforms

- data warehousing, GIS data integration and delivery to mobile devices
- human-computer interfaces with user acceptance

Examples will be given from design statements on the projects being undertaken at City University.

These factors reflect a technological view of the elements needed for the performance and functionality of LBS applications at a level of innovation likely to attract the consumer. The challenge for researchers is to demonstrate attractive potential applications, the challenge for the industry is to deliver the elements needed at a price the consumer and business users can accept. However, if either part in the equation is unrealistic, then LBS might not progress to achieve their potential.

1 Introduction

Delivering fully operational location-based services (LBS) is proving to be a difficult task. Only a small number of LBS are operational and revenue earning in 2006 if we define LBS as fully location-aware applications delivering services through client-server or peer to peer architectures over wireless, cellular and satellite communication networks. In one sense this is surprising given the huge interest in LBS, in another sense it seems that LBS are simply following a well worn path through the 'hype cycle' (Gartner 2005). In 2003-4 LBS were passing through the 'peak of inflated expectations’; while in 2005 they were in the 'Trough of disillusionment'. If LBS are to progress up the 'Slope of enlightenment' to the 'Plateau of productivity' an assessment is needed of what is needed for LBS to succeed.

This paper is focussed on what the experiences of research and development into LBS have told us about the design constraints on operational LBS. This work is going on in a range of disconnected communities e.g. ubiquitous computing, mobile (geo)web services, geopositioning, telecartography, mobile HCI and personal navigation to name just a few, and few researchers have attempted a synthesis (Lopez 2004). This paper tries to assess what is now known with confidence about LBS design, in order to encourage convergence and greater cooperation between the various players in the chain of LBS relationships required to make these systems work (LBS value chain, figure 1).

Fig. 1. LBS value chain (courtesy of C. de la Fuente)

Finally, specifying definitions is important in a rapidly developing field. The assumption is adopted here that the following are not LBS within the richer sense defined here:

- autonomous clients with positioning, locally stored content and limited network updates (e.g. in-car navigation tool with traffic updates)
- mobile devices with low resolution geopositioning delivered by mobile telephony such as cell global identity, and simple tiled service scopes, such that the relationship between location and scope is weak (e.g. where's my nearest functions on GPRS equipped phones)

This assumption exposes the author's view that LBS must be able to self-customise their services, based upon high resolution physical location and semantically rich location context, and that this is the definitive core functionality of LBS. This is not to say that web and WAP delivered geographic information is not valuable in itself, merely to emphasise the distinction between Google Local directions on a mobile device and an application like allsportgps.com.

2 Two LBS that work

LBS are a class of applications where there is a definitive design benefit from actually fully implementing the system in a real environment. Since LBS are distributed, componentised and dependent on a range of associated services such as positioning input, many LBS as designed would not perform acceptably in practice. There is a natural tendency of designers to expect software (re)engineering to solve these performance issues. However, in LBS the performance issues can only be solved by compromises on data models, positional update and user interfaces once the system as a whole can be evaluated in use. This means that effective design requires many cycles of (sometimes comprehensive) redesign in close communication with users, preferably involving users choosing to spend their own money in the later stages of development. Cost is not the only way to improve design, but since LBS are designed as personal services, and services need revenue streams to sustain them, it is one of the most effective drivers. In some cases it may be possible to demonstrate that a technically excellent system cannot be implemented at a price that users among the general public are prepared to pay! This is a real discipline for designers who are used to systems (e.g. GIS) designed for small numbers of (relatively) highly trained users in the professional market.

The following sections introduce some operational LBS to present lessons learned from their development.

2.1 Pedestrian navigation: KDDI EZ Naviwalk

KDDI are a major Japanese mobile phone operator with a dominantly 3G network infrastructure. KDDI operate a positioning service for developers based on gpsOne assisted GPS (A-GPS) technology from Qualcomm, which allows real time positional updates to be calculated on the mobile device after the network-assisted first fix. KDDI phones support GPS-aware applications written in BREW, Qualcomm's mobile device development language.

The EZ Naviwalk service developed by Navitime offers pedestrian navigation support to subscribers at the equivalent of €2 per month, plus the cost of the data package, which is usage-related. The service offers the following functions:

- display of the current position
- origin and destination routing (by use of bar code lookup from map books and magazine adverts) (figure 2)
- voice guidance on routing
- search around current position
- finding nearby stations and their train services
- auto-heading up while moving (with Kyocera A5502K GPS handset with electronic compass)
- ability to record current position in 'My spot' recording service
- Maps are downloaded to the device as needed (e.g. when a wrong turn is taken) and the destination can be changed *en route*.

EZ Naviwalk has over 400,000 subscribers on KDDI's network indicating a high level of level of acceptance amongst users. The service has been refined, with feedback from users, to have a fast autonomous positioning mode after the network-assisted first fix, and orientation support with the digital compass as this means that the map is always oriented the way that you hold the device.

Fig. 2. Sony Ericsson W21S phone showing EZ Naviwalk directions function

2.2 Mobile multimedia location guide: Camineo guides

Camineo is a spinoff company exploiting technology developed during the EU Information Society Technologies programme project 'Webpark'. Camineo Guides are location- sensitive multimedia guides for tourism delivered on mobile devices with A-GPS or Bluetooth GPS, web browser and java virtual machine, such as the HP iPAQ 6515 or QTEK (Kjeldskov, Graham et al. 2005).

Camineo guides are constructed on a web services model with:

- interfaces to position services such as connected GPS devices or A-GPS
- a mini 'web portal' (web server) on the mobile device itself, to serve web pages with multimedia content, and provides servlet support for java applications;
- a network interface to GPRS and UMTS cellular services to allow access to web services on the Internet/ intranets such as Web Feature Servers;

- java applications generic to the Camineo platform such as map viewer and geo-bookmarking tool (users can place timed and located bookmarks while using the service);
- java applications specifically written for the Camineo guide location (in the SNP these include a hiking progress monitor and animal/ plant species guide, both of which can take advantage of the current position;
- a capability to visualise or explore points/ areas of interest, paths and geo-bookmarks formatted in appropriate web formats and assembled in a database by thematic and geographic criteria;
- an advanced search facility that uses probabilistic search to predict the user's future position and returns results for points/ areas of interest, paths and geo-bookmarks based on current and/or predicted location;
- a web browser interface with a user-friendly 'my Guide' front page which can also run in kiosk mode with reduced buttons and menus.

Fig. 3. Camineo/Webpark front page

Camineo guides are being used by national parks such as the Swiss National Park (SNP) and visitor Centres like the Ecomare centre on Texel Island, Netherlands to provide interactive, location-sensitive guides for visitors. Devices are rented out to visitors at the equivalent of €5 per day in the SNP fully inclusive of network and content charges. All available devices have been fully booked for rental out over the summer 2005 season in the SNP after three season- long cycles of intensive development/ feedback.

3 What these working LBS show us we need in future

3.1 Concepts of context and activity

LBS are examples of scaleable ubiquitous computing applications designed for use in environments ranging from individual buildings to cities and even whole regions in the case of national parks. As such, LBS must operate with a sense of space and place for context (Raper 2000) through which it can deliver its location sensitivity. The sense of space must encompass a transformation between the location sensor referencing and the service referencing, for example that required by a map. In GPS-focussed services GPS referencing using latitude and longitude on the WGS84 datum must be transformed 'on the fly' to the local/ national coordinates used by the LBS (e.g. Swiss Grid for the Camineo SNP Guides).

Providing the sense of place is the more demanding function for LBS to achieve as this involves the cognitive domain of the user. There appear to be a spectrum of possible approaches, with place ontologies at one end (e.g. legally defined national and local government regions), which need to exhaust space in order to be effective. These approaches often have a poor match with the user's cognitive domain.

At the other end of the spectrum there are cognitive definitions of place defined by the user either passively or actively. These definitions should have a strong correlation with the user's cognitive domain and a complex relationship with the system-implemented sense of space. The Camineo guides have a system to forecast the likely location of the mobile user based on previous behaviour that uses probabilistic approaches to create a future location 'likelihood surface' that can be used to rank information resources (Mountain 2005). This technique is an attempt to model what the user may see as the area from which relevant information resources might be drawn. In geographic information retrieval terms this is the geographic relevance of a query. Raper (in prep) suggests that the identification of geographic relevance of a query may be the way forward in enriching the concepts of place in LBS.

Dransch (2005) has suggested that concepts of context need to be complemented by concepts of task and activity, for example through activity theory. Activity theory models “a conscious and directed act that is executed to reach a defined goal” (p33) i.e. a cognitive action plan, which is composed of goals, sub-goals and actions. Artefacts (such as LBS) are needed to help achieve goals (e.g. wayfinding), although they modify the action plan. A 'gulf of execution' has to be bridged if the artefact is to help, which requires a study of the mediation role of the artefact.

When analysed from this perspective, it is clear that LBS must be designed to help with the next goal in a way that users can accept and control. In the LBS4all project at City University, the user interface to a LBS for people with mobility problems (e.g. visually impaired and older people) based on the Camineo platform, is being evolved from an artefact focussed on wayfinding to one focussed on commentary. On the basis of fully-engaged user testing it is clear that wayfinding still needs to continue to depend on primary navigation tools e.g. the white cane for the blind. Meanwhile the LBS can 'commentate' on the surroundings to open up new opportunities for the user and enrich their knowledge of the physical and informational environment.

3.2 Context architectures

Concepts of context and activity need to be operationalised for rich LBS applications. Reichenbacher (2003) has defined (computable) context as a function of the following functions and attributes:

- user (user profile, identity, privacy)
- technology (device capabilities, network, transport protocol)
- activities (the locating, navigating, searching, identifying tools)
- location (mobile positioning, geocoding)
- time (system time, real time)
- information (sources available via WMS, WFS, gazeteers)
- interface (map/ information presentation format).

This context can be defined and maintained using these functions and attributes, although as yet we know little about how to characterise some of them (e.g. privacy) or how they should be combined and traded off (e.g. rich services in poor network environments.

José, Moreira et al. (2003) defined the AROUND context architecture to enable the discovery of LBS on the Internet through proximity models based on distance and service scope (where the service scope is the location context). Location contexts for service scope can be aggregated using spatial semantics, resulting in a graph structure. AROUND supports:

- containment, with many to many relationship allowed between contexts.
- adjacency, allowing contexts to find their neighbours, or to support pre-fetching to allow rapid context changes.

The AROUND architecture is implemented as a relation between current 'base' context and metrically or semantically close contexts mediated by a name service to resolve global location context names into references to specific AROUND servers at which they can be accessed using CORBA protocols.

An AROUND-type city guide client was implemented in java for a connected mobile device, operating in push mode as the user moved through the contexts as part of the Hypergeo project (a precursor project to Webpark). The chief challenges such architectures pose are the creation, maintenance and semantic relations of contexts in a service context with no relevant standards and poor scalability and security to build on. In the development of EZ Naviwalk the context can be based on distance or service scope (e.g. railway stations), although the cognitive domain is not complex.

3.3 Positional fusion and rationalisation

A wide range of sensor technologies, each with their own specific error characteristics, can provide geopositioning for LBS. Rizos (2005) divided the technologies into those that require additional installed infrastructure (such as WLAN/ UWB/ Bluetooth/ RFID technologies), those that can be delivered with existing mobile telephony infrastructures, and those ubiquitous positioning technologies such as GNSS and intertial navigation / dead reckoning systems. Only the latter (specifically GPS and A-GPS) currently provide the combination of service availability and accuracy that has been sufficient to offer geopositioning to revenue earning LBS beyond the simplest 'where's my nearest' services.

Hightower (2002) note that '...each ubiquitous computing system typically treats (location) data in its own idiosyncratic manner' (p22) and that there are lots of new location determination technologies available. This implies that location sensor (data) fusion is likely to become important. Solutions will either be:

- rigid vendor-integrated systems for specific applications, or
- robust software abstractions connecting multiple sensing technologies and multiple applications

To achieve the latter Hightower (2002) and Graumann (2003) proposed the 'Location Stack', modelled on 7-layer OSI networking model, with the following layers (from bottom up):

- sensors, exporting raw data values
- measurements, raw sensor data plus uncertainty related to the sensing technology
- fusion, of measurements, definition of geodetic frame
- arrangements, reasoning about location information produced
- contextual fusion, merging location and non-location data e.g. states
- activities, relating location to semantic states
- intentions, cognitive desires of users

To date, most LBS have been implemented using single geopositioning solutions to avoid the need to deal with location sensor (data) fusion. The challenges of integration of GNSS and intertial navigation (IN)/ dead reckoning (DR) systems in particular pose significant problems, since there are significant non-linearities and latencies involved in fusing geopositioning derived from time (GNSS) with those derived from space (IN/DR) methods.

3.4 Application development

LBS application development poses significant challenges in the light of the length and complexity of the LBS value chain (figure 1) and the need to integrate between the tiers of a client-server architecture. There are three basic elements to a tiered client-server architecture. Firstly, the data storage in a DBMS, secondly the application or 'business logic' (consisting of the communication handling and data processing), and thirdly the graphical presentation through a user interface (GUI) whether that be a standalone application, a Java applet or web browser. There is a spectrum of possible physical implementations ranging from all three elements running on one machine, to each of the three elements running on a separate machine. An 'n-tier' architecture partitions the logic or data onto multiple machines when this is required to handle (or specialise) the load. Most of the LBS solutions available use the client-server approach (Meng, Zipf et al. 2005), with the exception of multi-agent designs.

The web services architecture of the W3C is the most generic client-server framework available and has created a universal platform for distributed processing and data. This is the ideal architecture for an open, distributed and co-operating set of services such as those found in the LBS value chain. The advantages and disadvantages of the use of a web services approach for LBS are:

- data can be encoded in XML and GML, but this encoding lacks richer semantics than geometry;
- httpd can be used as a server protocol, but it is intrinsically stateless, making it difficult to maintain sessions;
- web services allow dynamic content to be served using Java servlets, but are not as efficient in transaction handling as CORBA;
- vector graphics can be delivered through SVG, but SVG coordinates are local not global;
- the user interface is a web browser

The Open Geospatial Consortium have defined Open Web Services for Geoservices and with OMA have defined the Open Location Services Geomobility server, both standard architectures for geospatial web services with associated information models allowing application development interoperability and convergence. Overall, the web services approach offers a flexible and standardised distributed application development environment, though it can deliver lower performance and has some semantic limitations in XML and GML. Camineo has used this approach to deliver its guides because standardisation offers guaranteed integration, longevity and the appeal of a familiar user interface.

By contrast KDDI have used a proprietary client server environment based around the gpsONE chip and the Qualcomm BREW development platform to deliver EZ Naviwalk. This environment offers performance and complete control over the user experience, but delivers the LBS through a custom interface and with hardware dependencies. This makes it harder to bring new developers to the platform, although this platform has already reached a considerable scale in Japan. An alternative proprietary LBS platform is the Drill-Down Server offered by Telcontar with its Rich Map Format, as used by Google Maps.

Both of the open and proprietary approaches are viable as demonstrated by the two applications profiled here, and each offers a set of advantages and disadvantages as part of their overall proposition. However, a key advantage of the web services approach is the speed with which it can adapt to the changes in the hardware, network environment and the mobile device operating systems by virtue of the open nature of the platform.

4 Conclusions

These factors reflect a conceptual, operational and commercial view of the elements needed for the performance and functionality of LBS applications at a level of innovation likely to attract the consumer. The challenge for researchers is to demonstrate attractive potential applications: the challenge for the industry is to deliver the elements needed at a price the consumer and business users can accept. However, if either part in the equation is unrealistic, then LBS might not progress to achieve their potential in the next few years.

Acknowledgements

Thanks to the LBS team at the giCentre at City University for input to this thinking, especially Vesna Brujic-Okretic, David Mountain, Fotis Liarokapis and Anil Gunesh. Thanks also to Clive de la Fuente for the LBS value chain diagram and much advice and assistance over the last several years. Work on the LBS4all project (http://www.lbs4all.org/) has been funded by the ESRC under grant RES-328-25-0007; work on the Locus project (http://www.locus.org.uk/) has been funded by the EPSRC under grant GR/T04212/01. The EU IST Programme funded the Hypergeo and Webpark (http://www.webparkservices.info/) projects. Grateful acknowledgement is given to these funders.

References

Camineo http://www.camineo.com/

De La Fuente C, Raper JF, Stanley M, Norris P, O'Donnell (2005) Location-based services: understanding the Japanese experience. Department of Trade and Industry. Globalwatch Mission Report. London, DTI. 68pp.

Dransch D (2005) Activity and Context - A Conceptual Framework for Mobile Geoservices. In: Map-based mobile services. L. Meng, A. Zipf and T. Reichenbacher (Eds.). Berlin, Springer: 31-41.

Graumann D, Lara W, Hightower J, Borriello G (2003) Real world implementation of the Location Stack: the Universal Location Framework. Fifth IEEE Workshop on Mobile Computing Systems and Applications, Monterey, California, IEEE.

Hightower J, Brumitt B, Borriello G (2002) The location stack: a layered model for location in ubiquitous computing. Fourth IEEE Workshop on Mobile Computing Systems and Applications, New York, USA, IEEE.

José R, Moreira A et al. (2003) The AROUND Architecture for Dynamic Location-Based Services. Mobile Networks and Applications 8(4): 377-387.

Kjeldskov J, Graham C et al. (2005) Evaluating the usability of a mobile guide: the influence of location, participants and resources. Behaviour and Information Technology 24(1): 51-65.

Lopez X (2004) Location-based services. Telegeoinformatics: location-based computing and services. H. Karimi and A. Hammad. Boca Raton, FA, USA, CRC Press**:** 171-188.

Meng L, Zipf A, Reichenbacher T (2005) Map-based mobile services. Berlin, Springer.

Mountain D (2005) Visualizing, querying and summarizing individual spatio-temporal behaviour. Exploring Geovisualization. J. Dykes, M.-J. Kraak and A. MacEachren. London, Elsevier.

Reichenbacher T (2003). Adaptive methods for mobile cartography. Proceedings of the 21st International Cartographic Conference, Durban, South Africa, International Cartographic Association.

Rizos C (2005) Trends in geopositioning for LBS, navigation and mapping. 4th International Symposium & Exhibition on Geoinformation 2005, Penang, Malaysia.

Location Based Services and GIS in Perspective[1]

Bin Jiang[1], Xiaobai Yao[2]

[1]Division of Geomatics
Dept. of Technology and Built Environment, University of Gävle, Sweden
Email: bin.jiang@hig.se
[2]Department of Geography, University of Georgia, Athens, USA
Email: xyao@uga.edu

Abstract

This chapter examines Location Based Services (LBS) from a broad perspective involving definitions, characteristics, and application prospects. We present an overview of LBS modeling regarding users, locations, contexts and data. The LBS modeling endeavors are cross-examined with a research agenda of geographic information science. Some core research themes are briefly identified.

1 Location Based Services: definitions, characteristics, and application prospects

With the rapid development and widespread deployment of information and telecommunication technologies integrated with lightweight mobile devices and terminals, pinpointing location on the move has become a common exercise. The technologies involve geographical information systems (GIS), global positioning systems (GPS), radio frequency

[1] This chapter is a reprint, with permission of Elsevier, of the same title initially appeared in the international journal Computers, Environment and Urban Systems, 30(6).

identification, and various other location sensing technologies with varying degrees of accuracy, coverage and cost of installation and maintenance. Some most recent location sensing technology based on ultrawideband radio can even achieve accuracies on the order of centimeters in an indoor environment. Meanwhile, the rapid evolution of cell phone industry from initial simple talk services to multiple functions of multimedia messaging and voice services with the emergence of broadband wireless infrastructure has created tremendous demands for various Location Based services (LBS).

What are LBS? There have been various definitions of LBS from different perspectives. One regards LBS as “any service or application that extends spatial information processing, or GIS capabilities, to end users via the Internet and/or wireless network” (Koeppel 2000), and another says that LBS are “geographically-oriented data and information services to users across mobile telecommunication networks” (Shiode et al. 2004). From a GIS perspective, the former definition concentrates on the GIS capabilities that are available in networked environments. The latter definition, on the other hand, narrows down specifically to geographic data and information services that are available in a mobile-networked environment. Both definitions emphasize that LBS are services targeted to a wide range of users. According to these definitions, both online map services (e.g. mapquest) and the Internet GIS can be considered important LBS applications, as they provide the kind of geographic information services via the Internet or mobile-networked environments to mobile devices. LBS are indeed partially evolved from the online map services and other Internet GIS applications, whereas current LBS mainly rely on lightweight mobile devices such as personal digital assistants (PDA), smart phones and wearable computers for delivering various services so as to provide added value to users. A true LBS application aims to provide personalised services to mobile users whose locations are in change. Location and context are the key players in LBS which are thereby often called location-ware computing or context-aware services.

Any definition of LBS would overlap partially with some key terms in research fields of GIS and geoinformatics. Instead of presenting a new definition, it is important to capture those distinct characteristics of LBS that differentiate them from other GIS applications. We can compare them in regard to five commonly accepted components of GIS, i.e. hardware, software, data, models, and people. In a comparison with conventional GIS, Karimi (2004) elaborated the distinct characteristics of LBS (he used another term “telegeoinformatics” to refer to LBS). From the hardware and software perspective, LBS are based on diverse platforms and packages which involve the use of Internet, GIS, location-aware devices, and telecommunication technologies. No conventional GIS applications involve so much diversity of hardware and software in an interoperating environment. In regard to data, LBS receive data from various sources such as remote sensing (including micro-sensors), positioning systems, topographic maps, and traffic and transportation data sources. The data from the various sources often need to be handled in LBS simultaneously and dynamically. Thereby LBS are much more heterogeneous in nature comparing to most other GIS applications. Because of various data sources involved, integrating the data and processing them in a real-time fashion seem to be more challenging. Moreover, models for generalization, visualization, and geoprocessing in general would also be imposed further research challenges because the user's locations are in constant change. Finally, human factors should be taken into account for any LBS application. Special considerations need to be taken for interface design, visualization methods, and reasoning approaches. More than often, user profiles and requirements need to obtain before and during any design and development.

Basic questions that LBS users are concerned about include: where am I currently? What and where are the nearest locations of interest? How to get there? The questions may arise in different contexts. LBS applications range a wide spectrum from daily life scenarios to specialized applications. A major application of LBS is to accurately position wireless emergence calls through E911 in the United States (or European equivalent E112) for emergency and rescue operations. Other applications include locating friends, locating nearest printer services, tracking staff, monitoring patients for emergency response, military training, asset tracking, and fleet management, to mention a few examples. The various information services can be delivered to the LBS devices in two different modes. The first is "push" mode where services are pushed to the user end automatically without the need of user request. The second is called the "pull" mode in which the user has to voluntarily request the information to be delivered from service centres. The market potential of LBS has been enormous, and thus it represents a new source of revenue opportunity. According to Allied Business Intelligence Research (http://www.abiresearch.com/), the worldwide market for LBS is to reach as high as $40 billion by 2006.

LBS not only can identify locations of human beings who carry location-aware devices, they also can track objects that are equipped with a tiny (and usually inexpensive) sensor identifier for delivering relevant services. For instance, products moving through the supply chain can be dynamically identified with embedded smart sensors, and massive products of such form a large-scale intelligent network (Swartz 2001). Such a sensor network has a better sense of customer's need and is able to deliver the related services intelligently. The sensor network described may sound very ambitious, but it represents some of current developments in pervasive computing. Pervasive computing (Satyanarayanan 2001) that reflects Weiser's initial vision on ubiquitous computing (Weiser 1991) represents the opposite of virtual reality, i.e. instead of letting people and objects immerse into the chips or computers, chips and microprocessors are embedded into human body and objects. This is likely to be a direction for future development of LBS. This chapter is not intended to go any further in regard to these future development directions. Instead, it strives to provide a perspective of the interplay between LBS and GIS in terms of modelling for LBS and to discuss research challenges that cut across the two fields.

2 Modeling for LBS

A good collection of models has been proposed in the past to capture important system components such as user, location, context, and data. These models are mostly introduced in a broader context of interaction systems of which LBS are an emerging type. The review that follows is based on the scholarly literature from relevant domains including ubiquitous computing, context-aware computing, general interaction systems, and GIS.

2.1 User needs and modeling

Users are central to LBS and so LBS applications should be designed based on a user-centered view. The user is a starting point for any LBS application design. It is the user who needs location-based services in various situations. User needs, user behaviors, and user profiles are important considerations in the course of designing LBS, since they determine what information should be provided and influence to a large extent the way systems and interfaces should be designed. User modeling sounds a very new subject for LBS, whereas it is an established domain in computer science for interaction systems (e.g. biennial conference series on user modeling since 1994). User modeling is referred to as "…the acquisition or exploitation of explicit, consultable models of either the human users of systems or the computational agents which constitute the system" (Csinger 1995, p. 32). This definition was given in the context of intent-based authoring that clearly reflects user's purpose of information presentation. Basic questions about the users in LBS are: who are the users? what are their needs? when and where do they need services? etc.. Many studies in user modeling (e.g. Jameson 2001) have examined a wide range of user properties including users' current states, behaviors and even long-term properties.

It is not an easy task to thoroughly understand the users and user needs, as they usually tend to be very diverse. Clustering the users in terms of interests, behaviors and personal profiles is an important step towards a better understanding of the users. For instance, to design a LBS application for a museum, the users can be classified in terms of the viewing habits and interests. Sparacino (2002) developed a wearable computer, so-called the museum wearable, to capture the user's behavior in visiting a museum. Three categories of the users can be identified, i.e. greedy users who need in-depth information on everything, selective users who want in-depth information on selective items, and busy users who see a little bit of everything. The information about users can also be captured by conventional user studies through personal interviews and field evaluations (Kaasinen 2003, Li 2005). It is important to note that user needs, interests, and behaviors are not static but rather in constant change. Information of such changes would be valuable for the design of an adaptive LBS system. Ashbrook and Starner (2003) introduced a model for predicting user's future locations based on the user's past locations. The model was verified by two scenarios involving a single user and multiple users respectively; refer also to Liu and Karimi (2005) for more recent advances on the issue. This kind of dynamic models would be highly expected in the future for developing LBS applications with a high level of intelligent responses and adaptation. The discussion of dynamics is probably more related to the next issue: location modeling.

2.2 Location modeling

If we unpack the term "location-based services", it is clear enough that location is an important part of LBS. Location is part of context (which will be further discussed in the following section) and it determines what information and services the user may expect. A location can be represented and perceived in different ways. A location could be represented as geometric or symbolic on the one hand (Leonhardt 1998), and it could also be absolute or relative on the other (Hightower and Borriello 2001). In GIS, locations are georeferenced in continuous or discrete georeferencing systems. For instance, a major entry of the University of Gävle is located at 17° 7' 9.23629"E, 60° 40' 7.53197"N using the universal reference system, known as World Geodetic System

1984 (WGS84). The location can also be represented as Kungsbäcksvägen 47, 801 76 Gävle, Sweden. The former is represented by coordinates in a continuous georeferencing system (WGS84 in this case) used by the GPS, while the latter is a visiting or postal address in a discrete georeferencing system. These two location representations are actually two georeferencing methods in GIS. Most indoor locations are represented in some local (rather than global) reference system. For instance, a robot can be located given a pair of coordinates relative to specific origin in a local reference system. The room numbered as 11:310 could indicate it is in the third floor of the building 11, whereas it is relative to the above university address.

Location modeling deals with the basic issue of representing space (or more precisely geographic space) for LBS. Two dominating methods in geographic representation are emerged from absolute and relative views of space, which arose from Newtonian and Leibnizian physics respectively. The former view regards space as a set of individual locations and objects, while the latter on how the individual locations and objects are interrelated within space. Two distinctive models: geometric and symbolic models (Leonhardt 1998) are de facto reflection of the above two views of space. The geometric models treat locations and objects as points, areas and volumes within a reference coordinate system, and they can support a range of queries regarding a position, nearest neighbors, and efficient paths among locations. Most existing GIS are actually based on the geometric models. However, the geometric models face some limitations and difficulties for the general public who are more used to linguistic expressions of spatial features, locations, and spatial relations. Research challenges in this regard will be discussed later. On the other hand, with the symbolic models, locations are modeled as sets and located-objects as members of sets, and interrelationships are established among a set of locations and a set of located objects.

Attempts have been made towards integration and extension of the geometric and symbolic models in order to take advantages of both. For instance, Leonhardt (1998) developed a semi-symbolic model in which a located object is represented as both absolute coordinates (as in geometric models) and memberships of named objects, i.e. symbols (as in symbolic models). Hu and Lee (2004) have recently developed a semantic location model that combines both geometric and symbolic aspects of locations based on location and exit hierarchies. A major advantage of this model is that it can automatically create location and exit hierarchy without human intervention. Along the same line of thoughts, we can remark that a street topology based on a graph theoretic representation (Jiang and Claramunt 2004) can support the kind of location modeling as well, since interconnection of named streets are clearly embedded in the modeling effort. The reader can refer to Becker and Duerr (2005) for a comprehensive overview of various location models.

2.3 Context modeling and adapting

Context is defined, for instance, as “location and the identity of nearby people and objects” (Schilit and Theimer 1994, quoted in Dourish 2004), or “location, identity, environment and time” (Ryan et al. 1997, quoted in Dourish 2004). So location is part of context, but context is far more than location (Schmidt et al. 1999). Dey (2001) defined context as “any information that can be used to characterize the situation of an entity,” where “an entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.” Context constitutes an important part, probably also the most difficult part of LBS, as both the user and location are part of the context. Because of the importance of context, LBS are often given other names, e.g. context-aware computing or context-aware services.

Context has impacts on information retrieval, user actions, and user behaviors with LBS applications. Contexts change persistently with mobile users, so context modeling must be able to capture the changes and reflect current context whenever and wherever the users are. Schmidt et al. (1999) introduced a model to better understand the concept of context for context-aware computing. The model adopts a context feature space, which is hierarchically organized. In the model, context is related to both human factors and physical environments surrounding the user. The human factors can be further subcategorized into user, social environment and task; the physical environments can be subdivided into conditions, infrastructure and location. The context related features can be further subcategorized as well. For instance, location could involve absolute position, relative position and co-location, etc. A similar consideration of context is made towards modeling people’s perception of distance in situated contexts (Yao and Thill 2005). To help context modeling or situation abstraction, a context toolkit was suggested for building context-aware applications (Dey 2001). Given the complex nature of contexts, it has been argued that an empirical and user-centered approach should be adopted to understand mobile contexts (Tamminen et al. 2004). In order to design mobile LBS systems that adapt to the changing contexts, contexts must be captured via sensor technology (Schmidt et al. 1999) or be taught through machine learning techniques (Laerhoven and Aidoo 2001). More recently Dourish (2004) presents an alternative model that focuses on a view of interaction rather than representation. It presents a new perspective towards a better understanding of contexts, although the model is not explicitly design- and technical-oriented.

2.4 Geospatial data processing and modeling

Geospatial data are one of the key components of LBS, as in essence LBS are a kind of data or information services. There has been increasing availability and continuous update of geospatial data over the past decades due to the advances in geospatial technology. Based on the projection of the National Research Council (2003) of the United States, the volume of geospatial data will increase by several orders of magnitude over the next decade. However the existing geospatial data infrastructure is not particularly suitable for LBS applications. For instance the data collected and maintained by the national mapping agencies do not match very well along the country boundaries. The data management and map symbols are not particularly designed for small mobile devices. All these impose challenges for data processing and modeling for LBS. Because of limited resources of mobile devices (e.g. limited size of screen and storage space), on-the-fly visualization and generalization are inevitable for mobile devices. In this respect, GiMoDig project (http://gimodig.fgi.fi/index.php) made efforts to develop various methods of delivering geospatial data and to provide data service infrastructure for LBS applications, although it is mainly limited to topographic data maintained by national mapping agencies.

While the existing geospatial data provide basic data layers, more data sources are needed depending on specific LBS applications. For instance, for LBS applications designed for tourist guide, all physical attractions including historic sights and shopping locations should be collected. Furthermore, some specific LBS applications have specific requirements for data modeling in some particular contexts. For instance it is found that landmarks are far more required than other information such as distance and street names (May et al. 2003) in a pedestrian navigation context. Major challenges for geospatial data processing and modeling include how to present information on a small screen in a clearly understandable way, and how to design maps adapting to changing contexts. No single map mode is for everyone, so multiple map modes are essential for various users. A visualization method that combines both 2D view and 3D view for wearable computers with navigation or wayfinding activities seems a solution (Suomela et al. 2003).

For the sake of convenience, we introduced the above modeling attempts from different perspectives involving the user, location, context, and data. Nevertheless, we must be aware of the fact that the considerations from all these perspectives should be integrated in the design and modeling process for any LBS application. For instance, prediction of a user's future location is discussed from both user modeling and location modeling perspectives. Both the user and locations are part of the context, therefore in essence they are inseparable in modeling processes. All in all, the four aspects should be coherently considered with a comprehensive conceptual modeling towards a systematic model for any LBS application.

3 Research challenges for LBS

The modeling attempts outlined in the above sections represent state-of-the-art research around LBS. The models are mainly proposed in the domain of ubiquitous computing with a few exceptions on geospatial data modeling conducted in GIS. LBS continue to be a hot topic in GIS, e.g. three short-term research priorities proposed by University Consortium of Geographic Information Science (UCGIS) are LBS related, i.e. LBS, social implications of LBS, and pervasive computing. In this section, we try to assess how the modeling issues briefed above actually constitute a series of long-term research challenges of geographic information science (GIScience). Table 1 lists the current ten UCGIS long-term research challenges (McMaster and Usery 2005). We shall first examine the connection between the research agenda and the modeling demands for LBS.

As LBS can be regarded as a special kind of geographic information services, it is no surprise that the UCGIS research agenda clearly has close links to the modeling issues for LBS. For instance, research on *spatial ontologies* with focus on ontological foundations for geographic information has at least two implications to the development of LBS applications. At one level, it can help to set up a common ontology for LBS for knowledge sharing among diverse users. At another level, it can help conceptualize design and modeling processes. Both location modeling and context modeling are related to the fundamental issue of *geographic representation* in GIScience. It concentrates on how geographic space should be represented conceptually and logically. The research issue is more challenging for LBS, because unlike other GIS applications where users' locations are not of particular concern, LBS are targeted to the users with constantly changing locations. Another UCGIS research priority, *spatial data acquisition and integration,* is also directly relevant to LBS. As a matter of fact, spatial data acquisition and integration is an integral part of data processing and modeling in LBS. Moreover, LBS applications often have unique requirements for data collection, integration, and accuracy analysis. Particularly, the issue of uncertainty of geographic information is closely linked to the data processing and modeling in LBS.

A more relevant topic is probably spatial cognition, which is inherited from long standing research interests in human environmental perception and cognition, map perception and interpretation, human spatial behavior, and wayfindings in complex built environments. The studies along this line can provide valuable inputs to the design and development of LBS in regard to human-environment interaction, human-map and -system interactions, user interface, and visualization methods. The challenge of visualization is closely linked to data modeling, and how geographic information is perceived, either via visual display or audio broadcasting. Due to the size constraint of mobile devices, graphic information should be represented in a simplified way but without loss of overall information. For the basic requirement, generalization linking to scale issue can help retain the simplified graphic forms. Furthermore research on space/time analysis and modeling could provide a powerful reasoning capability for more innovative value-added services. From a societal aspect, LBS are a key instrument for the improvement of the quality of life and personal productivity. On the other hand, societal impacts of LBS also include surveillance and invasion of personal privacy, and changes in human spatial behavior (Dobson and Fisher 2003).

Table 1. UCGIS long-term research challenges (McMaster and Usery 2005)

(1) Spatial ontologies
(2) Geographic representation
(3) Spatial data acquisition and integration
(4) Scale
(5) Spatial cognition
(6) Space and space/time analysis and modeling
(7) Uncertainty in geographic information
(8) Visualization
(9) GIS and society
(10) Geographic information engineering

Having elaborated on the close links of the research agenda of GIScience and modeling issues in LBS, we want to ask this question: what is special about LBS? Indeed, LBS represent some very special attributes of geospatial technology. For instance, most users of LBS are the general public; the user's behavior, location, and context are in constant change, and the systems must be adaptive to the changes. All these special characteristics are typically not dealt with in conventional GIS. In the rest of the section, we suggest a few research themes for future investigations, whereas the list of research themes is not intended to be exclusive.

3.1 Naïve users and next-generation GIS

A distinct characteristic of LBS is that they are generally oriented to naive users. Potentially everyone may become a user and therefore no assumption can be made about a user's prior knowledge of GIS or the spatial environments. While this fact provides a great opportunity for ubiquitous use of GIS, it also challenges GIS to cater for the particular needs of naive users. An average citizen usually has qualitative abstractions of the environment (Cohn and Hazarika 2001). Naive users tend to acquire commonsense, often qualitative, knowledge about the spatial structure of the geographical world through experiences without concentrated efforts. The knowledge may be incomplete or inaccurate at times, yet they still can be very powerful in making useful conclusions (Kuipers 2004). The commonsense geographical knowledge is usually expressed in linguistic terms such as place names (e.g. *Atlanta, White House, Main street,* etc.) and spatial relations (e.g. *north, in, near,* etc.). Therefore for the general public, the next-generation GIS not only should have the metric data handling capabilities, but also should be receptive to qualitative information and make best conclusions out of it.

The idea of next-generation GIS for naive users has close ties to several lines of intellectual investigations in the literature. Particularly, naive geography (Egenhofer and Mark 1995) provides the theoretical foundation for the next-generation GIS for naïve users. Naïve geography concerns with formal modeling of commonsense geographic world and the design of GIS for average citizens without major training in GIS or geography. A considerable amount of research has been made to contribute to naive geography. For example, Yao and Thill (2006) proposed a framework to handle locations referenced by qualitative spatial relations in GIS. Research achievements from a number of associated threads of investigation make direct or indirect contributions to naïve geography. These fields include qualitative spatial reasoning, perception and cognition of space, studies of the relationship between natural language and perceptual representation of space, computational models of spatial cognitive maps, uncertainties in spatial boundaries, as well as research on place names and digital gazatteer (e.g. http://www.alexandria.ucsb.edu/

gazetteer/). Researchers from computer science have also shown great interests in dealing with qualitative spatial information (e.g. McGranaghan 1993; Wang 2003).

In spite of the research efforts that have been made, current LBS and GIS applications are still in its quantitative-dominant stage. Coherent and consorted research is in great demand towards the realization of the next-generation GIS that cross between qualitative and quantitative paradigms. Research issues include mapping mechanisms between qualitative data and quantitative data, the design of user interface, interpreting semantics of linguistic expressions in LBS, incorporating qualitative spatial and temporal reasoning models in LBS, and visualization of qualitative location information, to name a few examples.

3.2 Spatio-temporal analysis and mining of mobile geospatial data

The geospatial data captured by mobile devices such as PDAs, mobile phones and wearable computers have been proliferating with the rapid development of LBS. This emerging data source has enormous potential to play an important role for our understanding of human activities and human behaviors in the environments. For instance, the Amsterdam RealTime project (http://www.waag.org/realtime/) collected massive data of individuals' whereabouts. From the datasets, researchers can track the spatio-temporal trajectories of the individual's activities. Spatial-temporal data mining algorithms can be used for the extraction of patterns from the datasets. An example of such patterns may be that many individuals in the study area go shopping on the way from work to home. Findings of human activity patterns and other spatial characteristics of human behavior can greatly facilitate the planning and decision-making processes, as these human activities and trajectories are sensitive to physical and cultural infrastructures (Ahas and Mark 2005). LBS provide a revolutionary data source of such data because they can collect spatio-temporal data that otherwise have to be obtained from very expensive data collection processes. This new data source will certainly stimulate more research towards what Miller (2005) called people-based GIS, with which spatio-temporal analysis and data mining is a major vehicle. Spatio-temporal modeling is likely to gain reviving research interests with the emergence of LBS.

Spatial-temporal analysis and data mining typically involve the use of vast amount of data and high computation load. Thus efficient data structures and algorithms need to be tailored for the LBS devices, which are typically not the top-of-the-notch computing environments in terms of storage volume or computation speed. In this regard, LBS data provide great opportunities as well as challenges for spatial data mining of human behavior data. Future research along this line include the development of data mining algorithms tailored for LBS data, design and implementation of data structures of activity-based location data, exploratory analysis of such data, and knowledge discovery from the data mining practices.

3.3 On-the-fly generalization and visualization

LBS can be characterized as map-centered geographic information services, as location information and services are most likely to be shown in mobile terminals. Conventional cartography, which was initially developed for stationary map displays, is rather insufficient to the particular needs of LBS. Distinct from a stationary cartographic system, the maps for LBS have various constraints such as small screens, persistent change of locations, and egocentric views. Special considerations should be given to these constraints for map rendering. For instance, the small screen constraint means that the conventional map with a great amount of details cannot be directly rendered for LBS applications. Mobile maps must be more simplified or generalized while retaining the necessary information. For this consideration, route maps or schematic maps (Agrawala and Stolte 2000, Avelar and Mueller 2000) are highly advantageous developments. Persistent change of locations implies the constant retrieval and update of the base map. Meanwhile, the retrieval and update must be adapted to the user's location and context. These issues have never been researched in conventional cartography, although the research issue of on-the-fly generalization has been explored for web mapping (e.g. Cecconi and Galanda 2002). In consideration of the abovementioned particular constraints and characteristics of LBS, more efficient algorithms are to be developed for on-the-fly and context-sensitive generalization and visualization.

It will also be interesting to investigate whether other visualization approaches, such as animation, multimedia, and multimodal geographical information presentation, are feasible for designing LBS applications. To adapt to the user's dynamic location, egocentric representations such as fish eye and variable scale maps, and panoramic view of the surrounding from the user's current position are appropriate. A recent book edited by Meng et al. (2004) presents a state of the art of the research and development along the line of map-based services. Many issues are still kept open for further research. These issues include human cognition of the new types of mobile maps, the usability of these maps, as well as the effectiveness of the mobile visualization methods. As the devices used for LBS are very compact, they typically do not include powerful input-output peripheries such as keyboards and mice. This brings about extra research challenges to facilitate human-device interactions that are necessary for advanced visualization methods.

3.4 Interoperability issues

Interoperability has been a challenging issue for GIS (Goodchild et al. 1999). Heterogeneity is also one important feature of LBS. The heterogeneity can be seen from various perspectives involving network protocols, hardware, software, positioning technologies, users, data sources and formats, and application semantics. Heterogeneity can be achieved through standardization. Currently many organizations are making contributions to the standardization of LBS to facilitate interoperability. Among them two bodies, Location Interoperability Forum (LIF) and Open Geospatial Consortium (OGC), have devoted significantly to location interoperability. LIF approaches location interoperability from the perspective of wireless network, while OGC targets the same issue from a geospatial angle. The two bodies endorse mutually some location interoperability standards. It is important to note that LBS are a collection of services offered by a value chain of interconnected companies from IT and geospatial industries. The companies include data providers, hardware and software providers, service providers, positioning data providers and so forth. The standards and open specifications significantly improve the efficiency of developing some LBS applications. In the future, more work needs to be done in order to achieve cross-standards interoperability.

For the development of LBS, it is rather important to ensure the meanings of concepts and data intended by the designer are effectively communicated from service suppliers to consumers (Kuhn 1996, cited from Raubal 2005). These are fundamental for semantic interoperability. Among the various types of interoperability, the semantic one is probably most difficult to achieve because of the linguistic sophistication and delicacy. It is challenging to catch and communicate semantic meanings across systems and among people. Raubal (2005) suggests adopt a people-oriented approach to solve semantic interoperability, as essentially meaning is not independent of people’s understanding and cognition. More research along the line is anticipated in the future.

3.5 Privacy and social issues

Although LBS indeed have enormous application potential for enhancing safety, convenience, and utility in our daily lives, people are wary of any abuse of the technology and location information. If not well guarded, LBS, like any other technology, may be reversed into the opposite of what was originally designed for (Sui 2004). LBS, as well as other location related geospatial technologies, threaten people and the society's privacy, or location related privacy to be more specific. The threat has been widely recognized and is vividly termed as "geoslavery" (Dobson and Fisher, 2003). The concern has enormous impacts on people's attitude towards using and adopting LBS. Current research on the issue of relations between the user's conceptions of privacy and intentions to use LBS is important for the development and deployment of LBS (Junglas and Spitzmuller 2005). Thirteen privacy issues related to the collection, retention, use and disclose of location information and technologies (Minch 2004) provide a full spectrum of understanding of location privacy. The privacy issues could be used as a foundation to build up a theory of LBS privacy as part of general theory of privacy in the information age (Moore 1997). Safeguard necessary for protecting rights of individuals must be provided to avoid abuse of location information and technologies, and to further facilitate the healthy development of LBS products and services.

Currently, research towards the privacy of LBS has been sporadic, among them most on conceptual and a few on empirical studies. For instance, the conceptualization of LBS as new media (Sui 2004) contributes to a complete and holistic perspective on LBS. In particular the detailed tetradic analysis of LBS based on McLuhan's laws of media provides some deep insights into social and spatial impacts of LBS on individuals and society as a whole. To alleviate users privacy fears, industry has started to implement some regulations to get rid of users' privacy concerns. The Privacy Management Code of Practice defined by Vodafone for instance allows the users to anonymize location requests by mapping the cell phone number to an alias (Spiekermann 2004). It also provides an interface option for the users to turn on or off localization. More research is needed on users' conception of privacy and how they shape their attitude towards LBS products and services. Future development of LBS and the fulfillment of their potential rely much on the advances around study, standards, and legislation about location related privacy.

4 Conclusion

This chapter strives to capture the current developments in LBS, an emerging and fast developing field cutting across the boundaries of geospatial, mobile, and information technologies. We have seen from the previous review that increasing efforts have been made by both geospatial scientists and computer scientists towards the advancement of LBS. We have also seen a series of issues and challenges imposed on LBS research from both technological and societal perspectives. The need and importance of many GIScience research topics find more justifications with LBS. Meanwhile these research topics also see new challenges with LBS. More cross-disciplinary endeavors are anticipated in the future particularly at the intersection of information technology, geospatial technology, and increasing awareness of social impacts of the technologies.

There is no clear-cut boundary of LBS and GIS, as many fundamental research issues of GIScience are those of LBS as well. The boundary could be even more blurry in the future when conventional GIS advances to invisible GIS in which GIS functionalities are embedded in tiny sensors and microprocessors. As speculated by Sui (2005), conventional GIS concepts may disappear, but instead GIS functionalities may appear in a pervasive fashion when the idea of ubiquitous computing comes true. The evolution of GIS concepts clearly reflects the shift of computing platforms from mainframe, to desktop, and nowadays to an increasingly pervasive fashion. It is the shift that makes LBS and GIS research special, challenging, and exciting.

Acknowledgements

We thank Daniel Z. Sui and Hassan A. Karimi, who read an early version of this chapter, for their constructive comments and suggestions. However, any errors and inadequacies of the chapter remain solely the responsibility of the authors.

References

Agrawala M., and Stolte C. (2000), A design and implementation for effective computer-generated route maps, *AAAI Symposium on Smart Graphics*, March 2000. 61-65.

Ahas R. and Mark U. (2005), Location-based services - new challenges for planning and public administration? *Futures*, 37, 547 – 561.

Ashbrook D. and Starner T. (2003), Using GPS to learn significant locations and predict movement across multiple users, *Personal and Ubiquitous Computing*, 7, 275 – 286.

Avelar S. and Mueller M (2000), Generating topologically correct schematic maps, *Proc. of the* 9^{th} *SDH*, Beijing 10-12 August 2000, IGU, pp. 4a.28 – 35.

Becker C. and Duerr F. (2005), On location models for ubiquitous computing, *Personal and Ubiquitous Computing*, 9, 20-31.

Cecconi, A. and Galanda, M. (2002), Adaptive zooming in web cartography, *Computer Graphics Forum*, 21(4), 787-799.

Cohn, A.G. and S.M. Hazarika. 2001. Qualitative spatial representation and reasoning: an overview, *Fundamental Informaticae,* 46(1-2), 1-29.

Csinger A. (1995), *Users models for intent-based authoring*, Dissertation, The University of British Columbia.

Dey A. K. (2001), Understanding and using context, *Personal and Ubiquitous Computing*, 5, 4 -7.

Dourish P. (2004), what we talk about when we talk about context, *Personal and ubiquitous computing*, 6, 19-30.

Dobson, J. E., and Fisher P. F. (2003), Geoslavery. *IEEE Technology and Society Magazine* 22(1), 47-52.

Egenhofer, M J., and Mark, D.M., (1995), Naïve Geography, In: A.U. Frank and W. Kuhn (Eds.), *COSIT `95 Proceedings: Spatial Information Theory: A Theoretical Basis for GIS*, Berlin: Springer, 1-15.

Goodchild M., Egenhofer M., Fegeas R. and Kootman C. (eds. 1999), *Interoperating Geographic Information Systems*, Kluwer Academic Publishers, Boston, MA.

Hightower J. and Borriello G. (2001), Location systems for ubiquitous computing, *IEEE Computer*, 33(8), 57 – 66.

Hu H. and Lee D. (2004), Semantic location modeling for location navigation in mobile environment, *Proceedings of the 2004 IEEE International conference on mobile data management (MDM'04).*

Jameson A. (2001), Modeling both the context and the user, *Personal and Ubiquitous Computing*, 5, 29 – 33.

Jiang B. and Claramunt C. (2004), Topological analysis of urban street networks, *Environment and Planning B: Planning and Design*, Pion Ltd., 31, 151- 162.

Junglas I. A. and Spitzmuller C. (2005), A research model for studying privacy concerns pertaining to location-based services, *Proceedings of the 38th Hawaii International Conference on System Sciences*, January 3-6, 2005, Big Island, Hawaii.

Kaasinen E. (2003), User needs for location-aware mobile services, *Personal and Ubiquitous Computing*, 7: 70 – 79.

Karimi H. A. (2004), Telegeoinformatics: current trends and future direction, in: Karimi H. A. and Hammad A. (eds.), *Telegeoinformatics: location-based computing and services*, CRC Press.

Koeppel I. (2000), *What are location services? - from a GIS Perspective*, ESRI white chapter.

Kuipers, B. (2004), Making sense of common sense knowledge, *Ubiquity*, 4(45). www.acm.org/ubiquity.

Kuhn, W. (1996) *Semantics of Geographic Information*. Department of Geoinformation, Vienna.

Laerhoven K. V. and Aidoo K. (2001), Teaching context to applications, *Personal and Ubiquitous Computing, 5,* 46 – 49.

Leonhardt U. (1998), *Supporting location-awareness in open distributed systems*, PhD thesis, Department of Computing, Imperial College London.

Li C. (2005), User preferences, information transactions and location-based services: a study of urban pedestrian wayfinding, *Computers, Environment and Urban Systems*, this special issue.

Liu X. and Karimi H. A. (2005), Location awareness through trajectory prediction, *Computers, Environment and Urban Systems*, this special issue.

May A. J., Ross T., Bayer S. H. and Tarkiainen M. J. (2003), Pedestrian navigation aids: information requirements and design implications, *Personal and Ubiquitous Computing*, 7, 331 – 338.

McGranaghan, M. (1993), Applications of Artificial Intelligence to Extracting and Refining Locality Information. In R. Fortuner (ed.) *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision,* Baltimore: Johns Hopkins University Press.

McMaster R. B. and Usery E. L. (2005), *A Research Agenda for Geographic Information Science*, CRC Press.

Meng L., Zipf A. and Reichenbacher T. (eds. 2004), *Map-based mobile services: theories, methods, and implementations*, Springer: Berlin.

Miller H. J. (2005) What about people in geographic information science? In: Fisher P. and Unwin D. (eds.) *Re-Presenting Geographic Information Systems*, John Wiley.

Minch R. P. (2004), Privacy issues in location-aware mobile devices, *Proceedings of the 37th Hawaii International Conference on System Sciences,* 5-8 January 2004 Big Island, Hawaii.

Moore, J. H. (1997), Towards a theory of privacy in the information Age, *Computers and Society*, 27(3), 27-32.

National Research Council (2003), *IT Roadmap to a geospatial future*, The National Academies Press, Washington, D. C.

Raubal M. (2005), Mappings for cognitive semantic interoperability, in: *AGILE 2005 - 8th AGILE Conference on Geographic Information Science*.

Ryan N, Pascoe J and Morse D (1997) Enhanced reality fieldwork: the context-aware archeological assistant. In: Gaffney V., van Leusen M., and Exxon S. (eds.) *Computer applications in archeology*, British Archaelogical Reports, Oxford, UK

Satyanarayanan M. (2001), Pervasive Computing: Vision and Challenges, *IEEE Personal Communications*, 8(4), 10 – 17.

Schilit B, Theimer M (1994) Disseminating active map information to mobile hosts, *IEEE Network* 8(5):22–32

Schmidt A., Beigl M., and Gellersen H. W. (1999), There is more to context than location, *Computers and Graphics*, 23(6), 893 – 901.

Shiode N., Li C, Batty M., Longley P. and Maguire D. (2004), The impact and penetration of location-based services, In: Karimi H. A. and Hammad A. (2004), *Telegeoinformatics: location-based computing and services*, CRC Press, 349 - 366.

Sparacino F. (2002), The museum wearable: real-time sensor-driven understanding of visitors' interests for personalized visually-augmented museum experiences, *Proceedings of Museums and the Web (MW 2002)*, Boston, MA, April 17 – 20, 2002.

Spiekermann S. (2004), General Aspects of Location-Based Services, In: Schiller J. and Voisard A. (eds. 2004), *Location-Based Services*, Morgan Kaufmann Publishers, 9 – 26.

Sui D. (2004), The Media and the Messages of Location-Based Services (LBS): death of distance or the revenge of geography? *Journal of Geographical Information Sciences*, CPGIS: Berkeley, 10(2), 166 - 174.

Sui D. (2005), Will ubicomp make GIS invisible, *Computers, Environment and Urban Systems*, 29(1), 361-367.

Suomela R., Roimela K. and Lehikoinen J. (2003), The evolution of perspective view in WalkMap, *Personal and Ubiquitous Computing*, 7, 249 – 262.

Swartz J. (2001), E-commerce and megamachines: identification, connectivity and inference engines, *Technology in Society*, 23, 159 – 175.

Tamminen S., Oulasvirta A., Toiskallio K., Kankainen A. (2004), Understanding mobile contexts, *Personal and Ubiquitous Computing*, 8, 135 – 143.

Wang, F., (2003), Handling Grammatical Errors, Ambiguity and Impreciseness in GIS Natural Language Queries. *Transactions in GIS*, 7, 103-121.

Weiser M. (1991), The Computer in the 21st Century, *Scientific American*, September, 1991.

Yao, X., and J. C. Thill (2006), Spatial queries with qualitative locations in spatial information systems, *CEUS*, 30(4), 485 – 502.

Yao, X., and J. C. Thill (2005), How far is too far? – a statistical approach to context-contingent proximity modeling. *Transactions in GIS,* 9(2), 157-178.

A Market and User View on LBS

Markus Uhlirz

Nokia Networks, Austria
markus.uhlirz@nokia.com

Abstract

This chapter briefly summarizes the options for the localization of a terminal, or cell phone, with current mobile networks. The market position of Location Based Services is then discussed from the point of view of a mobile operator's service portfolio. Finally, a user-perspective of LBS concludes this chapter.

1 Location Based Services

Since the introduction of mobile non-voice telecommunication services, the prospect of Location Based Services (LBS)[1] has driven engineering fantasies, seeking to combine the capabilities of a mobile communications device with real-time positioning information. Location-Based Services have been advertised and promoted as a "killer application" for third generation telecom networks. LBS has so far seen many false starts and even false prophecies concerning its immediate future. As with many other buzz-words, the concept of LBS has gone up and down the Telecom "hype cycle" [1]. While other telecom developments have reached a certain level of public visibility and commercial importance, LBS has remained at a

[1] In the absence of a generally accepted definition, the term "Location Based Services" (LBS) shall be understood here as "a location-sensitive or location-aware service delivered to users by means of a public mobile telecommunication system".

marginal level of user acceptance since its introduction and has not progressed in a noticeable way.

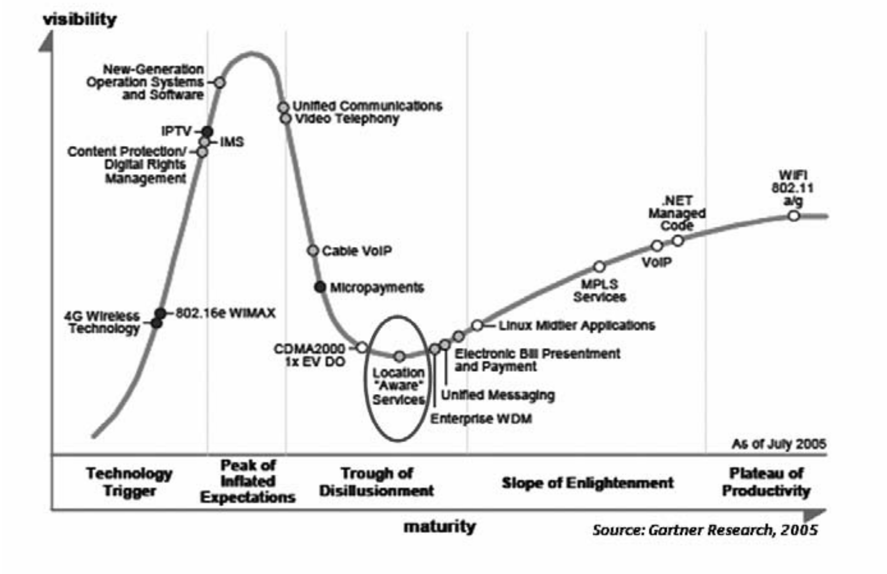


Fig. 1. Hype Cycle for TeleCommunication, 2005 (Gartner Research 2005)

1.1 Mobile Positioning and Location Based Service

Establishing the location of people and objects has become a commodity within recent years because of the public availability of the satellite-based GPS system [2] and cheap handheld positioning devices as consumer products [3]. Location-aware Services can be provided today with a GPS receiver and portable devices that acquire the necessary information from a static database integrated into the technical solution. Examples for such services are various map and routing systems as found in cars [4] or other handheld devices [5].

One main impetus for developing a telecom based positioning system was the requirement made by the Federal Communications Commission (FCC) in the USA in the late 1990s for a mobile phone based emergency roadside service that became known as "E911" [6]. As specified by the FCC, the E911 service shall allow positioning of a road accident with an accuracy of 500 feet (167m) in at least 95% of cases. This request has influenced many telecom and technology companies to investigate the feasibility of mobile positioning methods. Additionally, the potential of

LBS for Health services, public safety, professional logistics and personal life-style was seen as very promising. The revenue prospects of mobile advertising, better targetting of customers with promotional information based on their present locations has created great enthusiasm in the marketing world. A large variety of location-based services was identified and promoted accordingly.

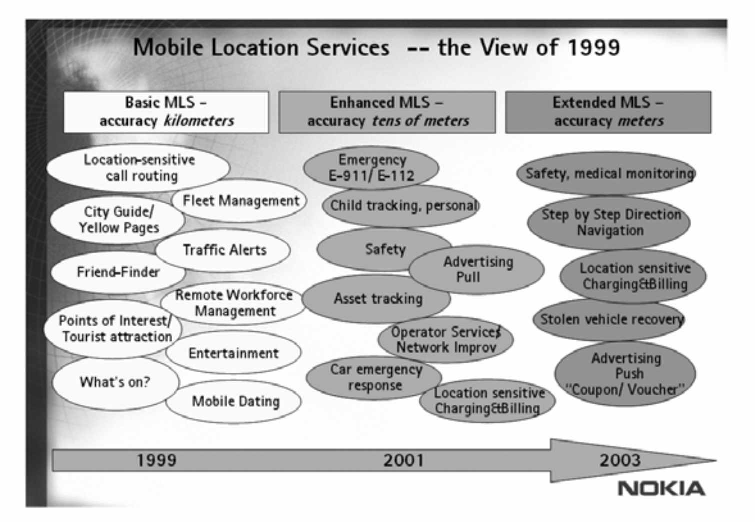


Fig. 2. Mobile Location Services as seen in 1999 (Source: Nokia)

1.2 Summary of Positioning Methods and Accuracies

GSM and UMTS systems were designed as telecom networks, in which accurate Mobile positioning was not foreseen. Due to the principle of distributed intelligence in the GSM architecture, there is not a single network instance that holds all necessary information to allow a accurate positioning of a device like a cell phone. This fact causes major difficulties for the implementation of LBS systems. Various mobile positioning methods have been suggested and developed. More detailed descriptions of the algorithms and applied methods are found in literature, e.g. in [7].

The main methods in use today are

- Cell-based positioning, reaching an accuracy of 0.5 to 10 sq.km, depending on the user's surrounding environment.
- Triangulation, based on signal timing differences as perceived by the terminal (E-OTD). This method provides accuracies on the order of a

few hundred meters and requires major investments to the network for the telecom operator.

- Assisted GPS: This method provides accuracies in the order of 10m, but requires either an built-in GPS receiver in the terminal or additional GPS receivers providing the assistance information. Very few high-end terminal models[2] support this type of positioning.
- Location Area Code: very coarse in resolution, typically 100 sq.km. and above. No new functionalities needed in the network. The method is not suitable for location-based services.

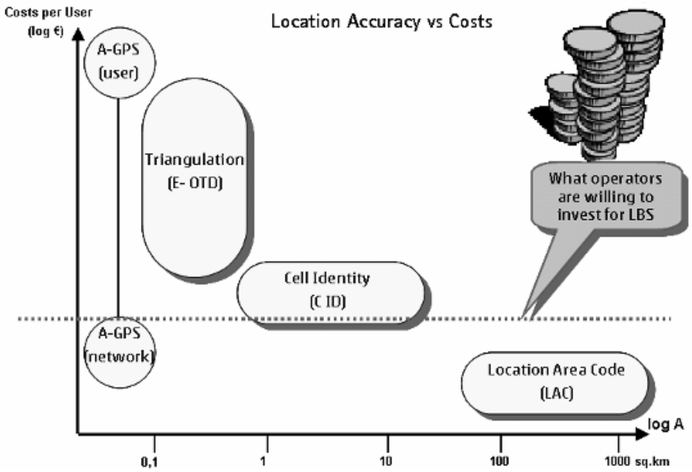


Fig. 3. Positioning Accuracy and Costs

The graph above shows the Costs per User against achievable area resolution for various positioning methods. The horizontal axis is area in logarithmic scale. The vertical axis is unscaled costs, but can be assumed to be also logarithmic.

2 Market aspects

2.1 Market trends

In the last years, market analysts and forecasts have been overly optimistic in estimating commercial importance and user acceptance levels of LBS. There has been a significant political pressure and the push to introduce LBS to the market in the “E-911” requirement in USA (1998/99), but

[2] e.g. Motorola A920, Nokia 7600 (both obsolete), Nokia N95 Communicator

industry has responded reluctantly in fulfilling these requirements. Today, E-911 is implemented only partially and is in various degrees of completion.

The supplier companies have re-focused their attention from broad consumer market to the corporate world, now providing specialized, tailored applications to a less price-sensitive audience. This move re-positions LBS from a mass-market product to a niche solution. Here it certainly has found its place and will continue to exist for a good while. Many of the anticipated mass-market applications envisioned in the late 90's have not materialized or have been withdrawn from the market. Winners in this development are manufacturers of internet-based mapping (Google maps, MapQuest etc.) and portable mapping solutions, which are becoming a commodity and a standard equipment also for cars in the Compact Class and medium price band (TomTom, Navigon etc.) Losers of this development are telecom carriers with own products and providers of purely telecom-based positioning systems.

The lesson to be learned from the story of LBS has only recently been made clear: Location information is an Enabler, but not a sellable service in itself. Location information is only valuable (hence sellable), if bundled with a service requested and desired by the user. Location information can very well be used to filter the relevant information elements from, for example, a comprehensive response to a Google query. This will immediately add value for the user by making the search results more relevant to his current situation. This is also a sellable item.

Looking towards today's telecom market, there are clear warning signals for LBS business:

- Unclear business model: No telecom operator has yet come up with a convincing business model for how to make money with LBS. Rather, location information is seen as a supplementary item to other information provided[3]. Currently, no mass-market service is known, that would fail if location information were not available.
- Marginal revenue streams: The net revenues from LBS continue to remain marginal. While data services in general are picking up momentum with introduction and acceptance of 3G services, LBS is being left behind. This discourages operators from further investments into the development of LBS, thus also creating a chicken-and-egg situation.

[3] Community- and photo-blogging portals and sites start adding location information to the displayed user contributions for purely informative or promotional reasons. Items can then be visualized on map-based web applications, e.g. Google Earth, google maps, yahoo maps etc.

- Not on Top-10-List: No operator worldwide lists LBS among his Top-10 data services. Several operators have already pulled the plug on LBS and have withdrawn it from their service portfolio, due to unfavourable ratio between costs and revenues.

2.2 Value Chain for LBS

The value chain for LBS is long and fragile, the relations between the different participating parties are complex and partly have competing goals. This makes a smooth cooperation difficult and may be one reason for the slow progress of LBS in the market.

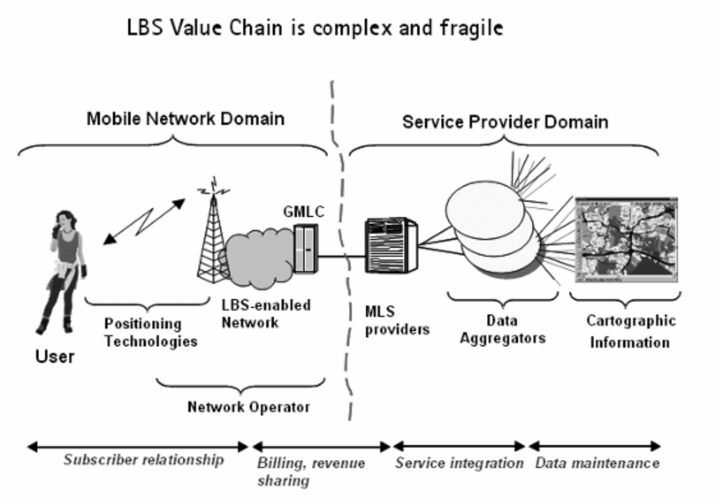


Fig. 4. Value Chain for LBS

The User is the requester and consumer of location-based services. Equipment for positioning technologies are supplied to the mobile network operator by various external parties[4]. The mobile operators provide their own branded portals[5] offering access to LBS services. The operator's Location Gateway is the borderline between the mobile Network Domain and the Service Provider Domain.

The Mobile Location Service (MLS) Providers are companies, specialized on providing customized location-aware software applications. They are provided with up-to-date and accurate geographical data by Data

[4] e.g. TrueLocation, SnapTrack, Nokia, Cambridge Positioning and others
[5] e.g. Vodafone Live!, O2 Active, 3 Geo and others

Aggregators, which in turn leverage their access to raw geographical information, such as detailed city maps, Point-of-Interest databases or electronic cartographic maps.

The Mobile Operator has a subscriber relation to the User, which is delicate: Subscribers are the operator's main source of revenues. They are costly to acquire and to maintain, easy to lose. The subscriber churn rate is one of the most closely watched business indicators for Operators. Therefore operators will be very cautious not to lose their subscriber's trust by compromising their privacy. Towards the other side the Mobile Operator has a fragile relation to the MLS provider. This relation is around fair billing and revenue sharing of the provided services as well as the trust, that the MLS provider will not disclose any user's personal information he may have access to in the provisioning of the services. On the other hand, MLS providers have an interest in gaining knowledge of subscriber profiling data and consumer's usage patterns, which they can in turn market towards other parties, e.g., direct marketing firms, advertising companies etc.

This conflict of interests has led to many legal discussions, delays of service launches and even to withdrawal of LBS offerings. As a measure of risk containment, some operators have imposed a "Code of Conduct" towards their MLS providers. The complications associated with privacy issues, the necessary investments for anonymising the LBS system, the overall too weak business case has caused several operators to reconsider their service offering and to withdraw from LBS all together.

2.3 Regulatory and Legal Aspects

Cultural difference in the valuation of personal privacy may explain the variation of progress speed for LBS in Asia, America and Europe. While Asian countries tend to be rather relaxed concerning user privacy, USA are more cautious and European countries take it very seriously. The European legislation has produced a number of laws and regulations ensuring user privacy. In July 2002, the European Commission has issued a Directive on Privacy and Electronic Communications [8], specifically addressing the issue of LBS. In particular, Art 13 ("Unsolicited Communications") of the Directive states that Users must not be mailed with marketing and other unsolicited communications without their prior explicit consent.

Furthermore, Art 9 ("Location Data other than Traffic Data") regulates that location data may only be collected from users after their explicit consent. Such data must not be stored any longer than required for the delivery of the requested service. Users must also be given by "simple means and free of charge" the possibility to temporarily refuse the further processing of their location-related data.

These regulations virtually rule out any plans around voucher distribution and location-based advertising, thereby denying operators access to a stream of revenues. It also requires the location process to include an explicit acknowledgement of positioning events by the User, making the process more complicated in handling. Mobile operators may be required to install an anonymising middleware solution in their system, in order not to disclose the User's personal data to any third party.

The given regulations and their financial impacts to the Business plan have caused the enthusiasm towards LBS to drop significantly with the mobile operators.

2.4 Ownership of Location Data & Privacy Issues

A number of legal cases have developed around the question of ownership of the User´s location data. Many third-party companies attempt to get hold of this knowledge as a re-sellable item. Mobile operators, however, knowing their fragile relation with their subscribers, do not want to take any risks with privacy issues and try to retain ownership of location data. The MLS suppliers claim, that the location information is extracted and made available by their applications and therefore claim the ownership rights. Some mobile operators have imposed a "Code of Conduct" with their MLS providers and intermediaries to ensure user's privacy on basis of these agreements.

Privacy aspects of LBS put operators into a dilemma:

- Privacy vs Ease of Use: more complicated handling of services
- Privacy vs Profit: high investments in privacy solutions needed
- Privacy vs 3rd Parties: disclosure of location data

These unresolved issues may explain why LBS have taken a rather slow start during the last years.

3 The User's View

3.1 Cartographic User interface

Cartographic display on mobile devices is a powerful method of conveying dense information content for those capable of reading maps. The challenges of rendering useful cartographic information on mobile devices are plenty and out of scope of this paper. Today's location-based services often do not necessarily require cartographic information or graphical map content output. Responses are frequently given in form of addresses and text-based routing descriptions. In cases requiring an accurate input of the user's position, the industry consensus seems to be, to provide the user with initial position estimation with an option of further manual accuracy refinement. This procedure can be map-based or text-based, using a selection list of e.g. street names or a free-text field for manual input. These methods are suitable for mostly urban areas and locations, in which external orientation in sufficient resolution is available and accessible for the user. The user experience of this procedure is poor. Users may abandon the transaction prematurely when confronted with cumbersome initial steps. A cartographic display can give very precise information and quick orientation. In urban areas, where routing distances are typically short, the precision of the initial position estimate needs to be significantly better than mobile positioning methods can currently offer. The information displayed to the user needs to be refreshed often to allow the user to monitor the progress of his travel on the screen. The continuous updating of the user´s position over a mobile network is a ressource-consuming and therefore costly task. However, users may expect a continuous moving map display as a feedback channel, similar to car navigation systems.

3.2 Suitability to the Use Case

A feasible, available, usable and cheap technology solution is not sufficient to make LBS a commercially successful service. As any other technology, LBS has a number of competitors with a very similar value proposition. To be successful, LBS needs to convincingly demonstrate its superiority to the User in the most common use cases. Some obvious competitors are the real-time experience, webpages, portable self-contained navigation systems on PDA's and CD-ROM media, just to name a few.

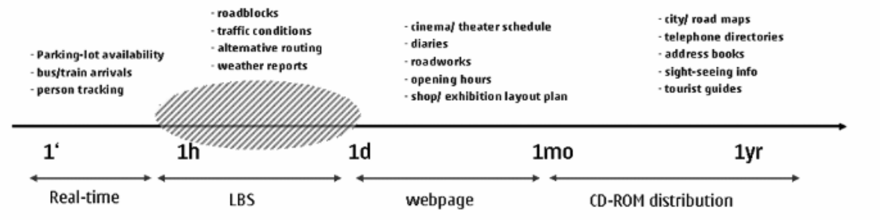


Fig. 5. Suitability of LBS on the time axis

Looking at the time dimension, LBS addresses only a small section of the time-axis. LBS can be considered best suitable in situations where information changes more frequently than once a day, but remains valid for duration of time in the order of travel-time needed to reach the destination. As an example, a “Parking-lot finder” is not a suitable application for LBS, because a parking lot’s average idle time until next occupancy in urban areas is substantially shorter than travel time needed to get there. In this case, the “real-time experience” is obviously the most suitable medium. Items that remain constant for several days or weeks are suitable for publication on web pages (event locations, cinema program, theater schedules, opening hours for pharmacies/shops, layout plans for temporary events e.g. trade fairs etc). Applying a LBS seems an inadequate effort for too little value.

Similarly, there is no real need to locate static objects via a LBS system. Street layouts, telephone directories or tourist guide-books do not change significantly in the order of months or years. Here, a CD-ROM edition makes good sense, while an LBS application to provide updates on changed street names and street positions seems rather inadequate.

LBS is most suitable for events without pre-determined locations and that remain valid within the order of several hours, e.g. traffic conditions, temporary roadblocks, weather report, rain radar, traffic cameras etc. This limits the scope of suitable applications to a rather narrow slice, but within this area, LBS can excel[6]. In the competition between static and mobile mapping, LBS will prevail when there is no other viable alternative to retrieve the requested information than via a mobile system.

A large portion of LBS’s lack of success is contributed by this inadequacy of service proposition and chosen presentation medium.

3.3 Critical Success Factors for LBS from User’s perspective

The critical success factor for LBS is the immediate value perceived by the User, irrespective of its technical complexity and actual precision of results. If an approximate result is perceived as “true”, users will use the service again. If a scientifically obtained, precise positioning result is perceived as “wrong”, the probability of a user re-trying the service is very

[6] Mapping company TomTom offers a supplementary feature for its Route planner product, allowing the user to access up-to-the-hour updates on traffic conditions and automatic re-routing of travel paths. This is a sensible and useful combination of pre-loaded map information on GPS receiver and incremental updates via mobile wireless LBS.

low. Based on the author's market observations during the recent years, the critical success factors for LBS include:

- Compelling need: The service must address a real need for users. "Nice-to-have" information is insufficient to sustain a business case. Location information is often seen as a complementary piece of information and readily accepted if free of charge, but seldom seen as deliverable per se.
- Immediate value to user: Location information must provide an immediately perceivable value to the user, then they are willing to spend money for this information and to use the service again. Too vague responses are unacceptable to the users and lower the interest in the service.
- Simplicity of use is a key issue. Currently, users rather refrain from using LBS if it requires manual data entries. Usage must be intuitive, results should be reachable within 3 clicks. Ideally, LBS would be voice-driven or produce results by automatic scanning/ reading of external info, such as bar codes, RFID chips etc. (e.g. voice in –> xml-script out –> input for mapping application)
- Quick processing time: Usability surveys for data services [9] show that users are impatient. Typically, after 6sec a majority of users tend to cancel the action. Out of these, half would retry, the others will abandon the attempt. A sand-glass as activity indicator apparently is not good enough to keep the interest awake. This also is a main restriction for amount of data that can be transferred to the user in a response.
- Near-100% hit rate in answers: Users are unforgiving: irrelevant or wrong answers are not tolerated. Users will accept answers preceived as "wrong" at most twice, then abandon the service, disappointed.
- Low or no price: Other than ringtones, wallpapers and music tracks, location information is a "non-storeable" item. It is consumed once, then it is worthless. Users are price-sensitive when payment is per event. Currently, the perceived value of location information is around 0.15 to 0.20 €/event[7]. Location information by itself has no commercial value and needs to be bundled with other items to become a sellable service.

LBS products are difficult to communicate. A high level of information abstraction and interpretation of results is needed by the User. Usability of services is not intuitive and requires too much manual input, refinement and validation.

[7] Test question: How much tip would you give a boy on the street to tell you where you are, when you are lost in a city?

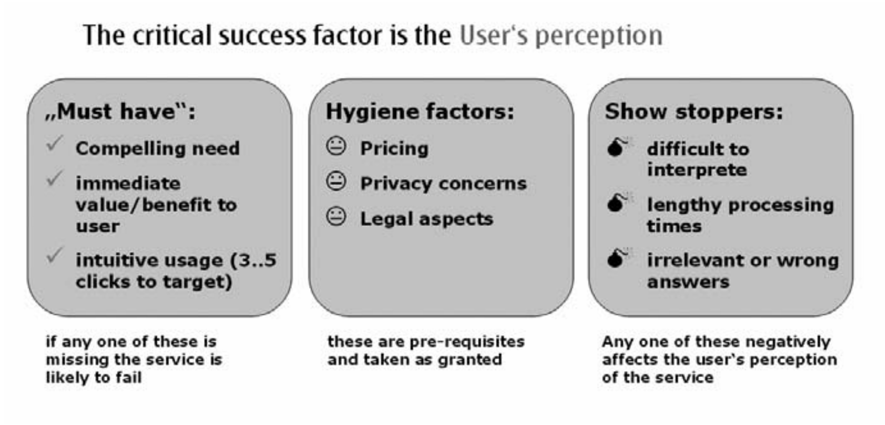


Fig. 6. Critical success factors

4 Conclusion

In this chapter we have summarized the available options for localization in a mobile network, assessed the position of LBS in today's service portfolio of a mobile operator's and have discussed possible reasons leading to the current market situation. Location-based services cover a small, specialised niche in today's telecom market. For a break-through, LBS yet needs to demonstrate its suitability for the user's everyday needs. Location-based products are difficult to communicate and to maintain. Usability of the user interface is still poor.

References

[1] Gartner Research, 2005.
[2] http://www.aero.org/publications/GPSPRIMER/
http://www.oreillynet.com/pub/a/wireless/2000/12/08/gps_intro.html
[3] www.garmin.com; www.magellan.com
[4] Blaupunkt Travelpilot: www.blaupunkt.de/
[5] www.tomtom.com; www.navman.com
[6] http://www.fcc.gov/911/enhanced/
[7] Retscher, G. (2002) Diskussion der Leistungsmerkmale von Systemen zur Positionsbestimmung mit Mobiltelefonen als Basis für LBS, in: Kelnhofer, F.et al (eds.): Geowissenschaftliche Mitteilungen, TUWien, 58, pp. 41-58, 2002.
[8] http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf
[9] Nokia Pulse Survey: Data Services Usability; Nokia & MORI, internal study (unpublished).

What makes Location-Based Services fail?

Gerhard Navratil, Eva Grum

TU Vienna, Institute for Geoinformation and Cartography
gerhard.navratil@tuwien.ac.at

Abstract

A lot of research projects deal with location-based services and also telephone providers services. Some services fail to become an economic success. In this paper we investigate possible pitfalls for location-based services. The categories of pitfalls are technical possibilities, legal restrictions, and usability. Using two example services we show how to determine the possible pitfalls.

1 Introduction

Discussions on location-based services (LBS) started in 2000 (Laurini 2000; Winter, Pontikakis et al. 2001). Since then a number of LBS were introduced by mobile operators. According to Berg Insight AB (2005) the revenues from LBS for Europe in 2004 was over 100 million Euro and they expect the number to grow by 2000% in 5 years. Not all LBS's are economically successful. The Austrian mobile operator ONE, for example, cut their LBS efforts in 2004.

Failure of a service can be viewed from two perspectives. The user defines failure as 'costs are higher than the benefit'. This includes, for example, information that is too expensive. The service provider defines failure as 'not profitable'. If the costs for the service are higher than the revenue (including indirect revenue like advertisement) the service is a failure. In this paper we discuss the problem from the view of a service provider.

Services create costs for the service provider and thus the service should produce benefit for the service provider to justify the expenses. Pitfalls may prohibit the benefit. Thus identifying pitfalls is important for developing location-based services.

Availability of data sets is influenced by three major aspects: Technical possibility, legal restrictions, usability (Navratil and Frank 2005). We assume that this structure does also apply for the success of LBS and the aspects, which let LBS fail, can be grouped into technology, legality, and usability and their interrelations.

In this paper we use two examples to show how to identify possible pitfalls. The examples are LBS, which do not yet exist but could be implemented. We assume that the LBS is designed for mobile phones only since according to Rümenapp and Gutsche by 2015 98% of the population in Germany between 15 and 55 years will own a mobile phone (Rüpemapp and Gutsche 2003). Other European countries will have similar mobile phone penetration.

We discuss the examples from the different perspectives defined above. In the first step we look at the technical status. The next step refers to limitations imposed by the Austrian law. Finally, we look at the usability. In the end we discuss interrelations between these aspects. This discussion shows that the aspects are often linked together.

2 Examples

Two examples shall provide the framework for this paper. The first example shows how LBS can be used to improve the international emergency call 122. The second example deals with pedestrian navigation in a very specific case: navigation at dealer shows.

2.1 LBS for emergency calls

The phone number 112 without any prefix is a European emergency call. 47 countries are involved in this program (European Union 2005). A call to 122 is routed to the closest emergency operator. The caller is connected to the emergency services in the area like police, ambulance or the fire department.

The operator needs information about the incident - what happened and where. Depending on the incident they will react in one of two possible ways:

1. If the incident requires help at the location the operator will contact according local emergency service and send them to the place of incident. The local emergency service will send one or more units to the place of incident and therefore need the address of the location or a route description. This information must be provided by the user (caller) and is forwarded by the operator. In some areas specifying a location or providing a route description is difficult, e.g., if a climber has an accident on a mountain. Route descriptions in rural areas are different than those in cities. There are no street names and house numbers, which can be used to describe the route. Thus different concepts like easily recognizable trees, rock formations, or clearings have to be used to describe the route.
2. It may also be that the user needs the location for the next office of an emergency service, e.g., the police. The operator must know the position of the user and can then provide a route description to the next police office.

Incidents usually require quick reactions and thus it would be helpful if the operator knew the approximate position of the user immediately. Then the operator can ask more specific questions and the system can automatically support him with additional information like, for example, a topographic map.

2.2 LBS for dealer shows

Dealer shows are an opportunity for companies to present their new products to potential customers. This requires a booth large enough to present the products. The corridors between the booths are kept as narrow as possible to fit the largest possible number of booths in the area available for the dealer show. The result is a situation where a large number of booths is situated in a small area and visitors gathering around booths block the corridors. Moving along the corridors is difficult.

Visitors face a problem if they want to find the booth of a specific company. Walking in a crowded area makes navigation more difficult. Looking at booth maps in a crowded area may even be impossible if the map is too large. Another problem with such maps is finding the current position by matching the visible companies with companies shown in the map. This is also a drawback for companies if they cannot be found. Thus an easy to use system for navigation would benefit both visitors and companies as we learned in discussions with journalists (Fürnkranz 2005, Katz 2005).

A navigation system for dealer shows can adapt results from pedestrian navigation services (Gartner, Frank et al. 2004). The user sends an SMS (short message service) to the service containing the name of the company he is looking for. The user then receives a description of the path to the booth of that company. This description may be either a unique message or a continuous stream of information. In the first case the quality of the description must be high enough to prevent errors and in the later case the service will have to track the mobile phone of the user and give directions when necessary.

3 Technical Solutions and their Pitfalls

Both examples comprise three steps. In the beginning the user accesses the service. Then the service processes the inquiry. Finally, the service provides information to the user. There are differences in the processing methodology of the two examples but the steps are the same.

A major difference between the two services is that the emergency call must not fail. Emergency services must be accessible anytime. Since emergency calls shall save lives considerations of costs are limited. The service for dealer shows should make profit and thus price and reliability of the solution must be balanced.

3.1 Access to the Service

The user must access the service. The access method for the emergency service is calling from a mobile phone. Depending on the rough location of the user the call is directed to an emergency response centre where a human operator receives the call.

The service for a dealer show must use an automatic method. Having a response centre in addition to the information desk, which is usually available, is not useful. Since the service shall relieve the information desk, connecting the service to the information desk is not wise. Thus, the access must use a method provided by mobile phones. Possible solutions are

- sending an SMS or
- using a WAP-service (wireless application protocol).

The user must provides the following information:

- Identification of the user for locating, response and payment
- Name of the company to be found

The telephone number provides unique identification of the mobile phone. The assumption is that the momentary user of the mobile phone is also the person registered at the telephone company. Problems with the identification may occur with pre-paid telephones because there is not necessarily a name and an address connected to them and thus billing may be difficult. Special treatment of such mobile phones may be necessary.

Another problem is the spelling of the company name. The service could provide a list of possible names if no perfect match is found. The user then corrects the name and again accesses the service.

3.2 Processing the Inquiry

A human operator is necessary to process incoming calls for the emergency call while the dealer show service requires an automatic response. The operator for the emergency call must provide help fast and efficient. The LBS can support him by supplying useful information. There is a list of questions to be asked by the operator to provide the following information:

- What happened?
- Where did it happen?
- Who is calling?

The system can automatically provide hints for the last two questions. The system can identify the owner of the mobile phone. This information can help reduce response times assuming that in general the caller will be the owner of the mobile phone. The system can also provide a position of the mobile phone and automatically come up with a map of that area. The operator can concentrate on the first question.

Maps can be provided in different ways. The easiest method would be using a web-map-service as provided by mapping agencies. Unfortunately response times depend on the workload of the web server used. A breakdown of this server could lead to a situation where the operator has no other information than the one provided by the caller. This is fatal if the operators are trained to follow a specific procedure, which depends on the availability of the map. A faster and more reliable solution would be to set up such a server within the local computer network. The agency then has full access to the server and can take precautions that backup servers are available. However, this solution requires more personnel and the emergency agency must pay attention to the quality of the data used, i.e., the agency must purchase and install data updates.

The operator finally has to decide, if he sends an emergency unit (ambulance, police, etc.) to the location of the user or if he guides the user to a location where he can get help. Depending on the situation only one of those possibilities may be applicable. In both cases the operator must provide instructions. Emergency units require less information than the users of the service, since emergency vehicles can be equipped with navigation systems. A route description for the user must consider that the user may be unfamiliar with the surroundings and thus needs detailed (and maybe even redundant) information.

In case of the dealer show the system must determine the current position of the user and the path to the position of the company’s booth. The current position can be detected automatically (e.g., by the cell ID or a GPS receiver within the mobile phone or connected to it). The quality of the detection (Retscher and Thienelt 2004) must fit the requirements of the service.

The definition of the path should be done using a suitable algorithm. A simple solution following the Dijkstra-algorithm (Dijkstra 1959) will not be sufficient because the path may become complicated. Different additions have been proposed by Duckham (2003), Grum (2005) and others (see for example Winter 2001; Beer 2002; Winter 2002).

3.3 Transfer of Instructions

The result of the service must be transferred to the user. The transfer of the location of incident to an emergency unit can be done as usual. A map of the area as used by the operator can be sent if necessary. Sending the path description to the mobile phone of a user is more difficult. The description must use landmarks and street names to describe the path. The transfer can then be done either as in form of an SMS for short paths or as a multimedia service (MMS). The advantage of SMS is that all mobile phones can process it. Older mobile phones may not have the capabilities to process an MMS. The disadvantage of SMS is the size limitation and the restriction to text. SMS do not allow images or maps.

4 Legal Restrictions

Legal restrictions for LBS originate from different parts of the law. Data is protected by copyright, programs may be subject to patent law (compare the discussion on software patents), and the user himself has privacy rights. These influences dictate procedures and may result in additional costs, e.g., for licenses.

The emergency call provides a more complicated case for a legal assessment than the dealer show. As said in 2.1 the emergency call works in 47 countries. Each of these countries has different laws. Some legal problems may only occur in some countries but not in others.

4.1 Access to the Service

Accessing the service is not restricted as long as there is no hidden data transfer to the service provider. An automatic download of the user's telephone directory, for example, would be illegal. Data transfer is necessary in different cases but it must be clear for the user which data is transferred.

It is also legal to charge higher rates than for normal telephone calls. However, the rates must be visible for the user. Since the emergency call is free of charge, this only applies to the dealer show example.

4.2 Processing the Inquiry

Both examples include locating the mobile phone. In case of the emergency call the user must tell the operator where he is. As said in 3.2 processing of the emergency call can be accelerated by automatically detecting the location of the mobile phone. The problem with this solution is that the mobile operator is not allowed to track the mobile phone. This includes storing previous positions as well as passing the positions to third parties. The service provider needs the permission of the user to do this. In traditional solution the user must agree that his telephone is located. The agreement specifies who may receive the data and how long it is stored. The agreement can either be sent by SMS or as a signed form by mail. This is possible for the dealer show but is inappropriate for the emergency call. There are two possible solutions:

- Users of the emergency call need help. The help either includes sending emergency units to the location of the user or telling the user where he has to go to find help. In both cases the user must provide his current position. Thus automatically sending the location of the mobile phone to the emergency service may be legal.

  The fact that emergency services shall save lives and thus the processing should be as fast as possible supports this hypothesis. As described in

3.2 the automatic location of the mobile phone helps the operator to react more accurate.

- Users must orally agree to locate the mobile phone. This could be done at the beginning of the call and recording the call provides proof, if necessary.

Processing the request requires data and software. Data are protected by copyright law. The service provider must have the right to use the data. The result of the process may be a map, which is transferred to the user or third parties (e.g. emergency units). This must be covered by the contract. The same is valid for programs. Programs written specifically for the service provider must be checked for collisions with patents. Software patents, as discussed in the European Union, may have a large influence in this area. Violations of patents or copyright laws may lead to lawsuits and, as a result, fines. Companies like Google register a large number of patents on software to avoid paying for other companies patents. They exchange their patents free of charge (Henzinger 2005). However, this is only possible, if a company develops software or algorithms.

4.3 Transfer of Instructions

Also during the transfer of instructions conflicts with patents and copyright law may occur. Transfer protocols and file formats may be protected. As said above also maps are protected by copyright law and using them (e.g., as a background) requires a license.

5 Usability

Usability has a different status in the two examples. The emergency call is free of charge. Simplicity and availability must be achieved at all costs to save lives. Finding the right way at a dealer show is convenience the user has to pay for. Problems with the speed of reaction may annoy the user but will not be as critical as in the emergency example.

5.1 Access to the Service

Accessing the emergency call should be as simple as possible. The steps of the process must be straightforward. The user should not get confused by tape instructions. A typical user of an emergency calls is nervous because he wants to report an incident and may be injured. Thus calming the caller is important. This can only be achieved by human operators who ask the user for all necessary data.

Accessing the service for the dealer show is different. Although the user will be in a crowded area, he will be less nervous than the user of the emergency service. Here the user wants to know two things before accessing the service:

- How do I access the service and how do I provide my destination?
- How much does the service cost me?

The user must also be capable of performing the task. The user must know how to write an SMS if this is the method to access the service. The second piece of information determines, if the user accesses the service at all. The user will not use the service, if the costs are higher than his benefits. This also includes the question, if the costs will rise, if the user misspells the name of the company. Since the benefit does not change costs should not rise.

5.2 Processing the Inquiry

Usability for processing emergency calls must be discussed from two perspectives: The user and the operator. The user wants to get help as fast as possible. Thus usability for the user is a measure for the response time and the capability of the service to adjust to user needs. The operator, on the other hand, defines usability in a more technical term: How much support does he get from the system?

The users define the quality of the service mainly by response time. The calls from February 23rd 2005 to April 21st 2005 in the hotline of the ambulance call center of Lower Austria contained (Bachinger 2005)

- 21 positive response and
- 115 negative Responses.

The negative responses covered, among others, the following topics directly related to response time:

- Patient died due to late response (close examination showed no evidence for a connection between the call center efficiency and the death of the patient)
- Unnecessary questions asked
- Ambulance did not find the way

European emergency call will be confronted with similar complaints. Especially the problems of finding the way and asking for necessary information will occur. The complexity of the European system is higher than that of the Austrian system because more types of incidents must be handled. The operator has more possibilities to react and must ask questions to decide on one of the possibilities. The sequence of questions is important to provide help fast. Also language problems may occur, if the user speaks a different language than the operator.

The operators need efficient methods for data collection and decision of further actions to minimize the response time. Automatic processes can support the operator. An example is the automatic detection of the user’s position. The system then can present the correct section of a map. The system can also support the operator when transferring necessary information to the necessary emergency organizations like police, ambulance, or fire brigade.

Usability of the LBS for the dealer show must only consider the user’s perspective. The user has a simple definition of usability: Does the service provide all necessary information to reach the booth of the company and does the service provide this information in an acceptable time span? Usability of the instructions will be discussed in the next section. The important aspect here is the amount of time between requesting the information from the service and receiving the instructions. Users will find the system inconvenient, if the delay is too long. Response times for information systems should be a few seconds when sitting in front of the system. Users of an LBS also want the answer in reasonable time. This limits the amount of time the system has to process the inquiry. The response should be faster than queuing up at the information desk. Users will be uncertain, if their request reached the system when the delay is too long and will try different methods to obtain the information.

5.3 Transfer of Instructions

The result of the service is a path description for the user or emergency vehicles. Only the emergency service may result in a path description for an emergency vehicle. A path description for the user is the result for the dealer show service but it can also occur in case of the emergency call service. The requirements for these two cases are different.

Guiding emergency vehicles to a specific location requires unambiguous identification of the location. Coordinates in a standard coordinate system may provide this reference. In rare cases additional information may be necessary, for example, the floor number, if the location is within a building. Problems may occur, if the location of the incident is a tunnel. Then the emergency organizations may be more interested in the entrance position to the tunnel and a route description within the tunnel. In general however, a single set of coordinates will be sufficient and navigation systems within the vehicles then lead the vehicle to the location of the incident.

Providing route descriptions for the user cannot assume the existence of a navigation system. Maps are difficult to read on the screens of mobile phones due to the limited resolution of mobile phones. Textual descriptions (e.g., lists of instructions) may be easier to understand but it is more difficult to create them automatically. In addition they usually provide less feedback for the user because there is only limited redundant information. Redundant information is necessary to ensure the user that he is still on the correct path.

All types of path descriptions should include landmarks because they deal with two problems: They allow setting the starting direction (‘move towards the traffic light’ is usually more helpful than ‘move south’) and they provide feedback. Raubal and Winter (2002) showed how to use landmarks in path descriptions. The selection of landmarks for route descriptions has been described in literature (Winter 2003; Nothegger, Winter et al. 2004). Logos of companies along the route provide landmarks for the dealer show example. The logos will be clearly visible since they shall attract potential customers. Since they are also distinct from each other they can be used as landmarks.

6 Combined Influences

Until now we discussed the influences independent of each other. Sometimes the influences are interrelated. The technical solution may, for example, have an influence on the usability of the service or may include some problematic legal aspects. In this section we will discuss the relations between the influences.

6.1 Technology – Usability

A combination of technical solution and usability provides the cost balance. The technical solution determines the costs. The provider of the service must pay for development of the service, acquisition of data, and running the system. The costs for each of these parts are determined by the technical solution. The users pay for the service creating income for the provider. The usability determines the number of users. If, for example, the user cannot determine the costs of the service in advance, he will not use the service. Thus the providers shall aim at technical solutions that include low costs and provide usable systems. This connection is usually discussed in a feasibility study.

6.2 Technology – Legality

Not all is legal that is technically possible. It is possible for the provider to locate a mobile phone. Locating the position of the mobile phone in relation to the network of the mobile operator is even necessary, if the provider must establish a connection (someone is calling). It is not allowed to store the location or to transfer the location to a third party. This creates problems if the LBS is provided by a company different from the mobile operator since that company will not have data on the users position. The solution must thus not only deal with the technical process of getting the location from the telephone provider, but also with getting the agreement from the owner of the telephone.

Another problem is using data formats or algorithms. The technical solution may ask for specific formats or algorithms. Relying on specific technology may create legal problems with software patents. Acquisition of licenses for a patent results in additional costs that have to be considered.

6.3 Legality – Usability

Sometimes specific treatment may be convenient for the user. Fleet management requires information on the position of the fleet members. A simple way to determine the position is locating the mobile phones and using their position. As discussed in 4.2 the mobile operator is not allowed to transfer the position to a third party. The only way to provide this information is letting the users of the mobile phone agree to the transfer. In Austria this is even necessary if the company using the fleet management is the owner of the mobile phones.

7 Discussion and Conclusions

In this paper we showed that technical solutions, legal restrictions and usability influence the design and efficiency of LBS. Some mobile operators offered LBS but failed to get a positive financial return from the service. The important question is why some LBS work and others do not.

A reason for failure could be that one of the three influences limits the service. It may be that the technology simply does not allow locating the mobile phone accurately enough or the LBS is not accepted because it is too difficult to use. Also, threats of a lawsuit for violation of patents or copyright law may stop an LBS.

When designing an LBS we thus have to answer three questions: Is this technically possible to provide the information? Is it legal to provide the information? It is possible to provide the information in a usable way? If the answers to all three questions are positive, we can continue with the process of providing the service. The topic then will shift to more economic questions, such as “Will we have more income than costs?” However, we must first make sure, that the LBS is possible, legal, and usable. In the chapter, we showed with two examples how the questions can be addressed to assess the potential and possible pitfalls of the service.

References

Bachinger, G. (2005). NÖ Rettungswesen Hotline - Endbericht. St. Pölten, Austria, NÖ Patienten- und Pflegeanwaltschaft: 10.

Beer, S. (2002). Wegfindung mit strukturierter Routenbeschreibung. Institute for Geoinformation. Vienna, Technical University Vienna.

Berg Insight AB (2005). The Structure Of The European LBS Market 2005. Gothenburg, Sweden.

Dijkstra, E. W. (1959). "A note on two problems in connection with graphs." Numerische Mathematik(1): 269-271.

Duckham, M. (2003). Simplest Path: Automated Route Selection for Navigation. COSIT 2003, Kartause Ittingen, Switzerland.

European Union (2005). http://www.sos112.info/.

Fürnkranz, G. (2005). Personal Communication. Businessword Magazine.

Gartner, G., A. U. Frank and G. Retscher (2004). Pedestrian Navigation System in Mixed Indoor/Outdoor Environment – The NAVIO Project. CORP 2004 & Geomultimedia04, Vienna, Austria, Selbstverlag des Institutes für EDV-gestützte Methoden in Architektur und Raumplanung.

Grum, E. (2005). Danger of Getting Lost: Optimize a Path to Minimize Risk. CORP 2005 & Geomultimedia05, Vienna, Austria, Selbstverlag des Institutes für EDV-gestützte Methoden in Architektur und Raumplanung.

Henzinger, M. (2005). The Past, Present and Future of Web Information Retrieval. Vienna, Austria: Discussion after the presentation.

Katz, F. (2005). Personal Communication. HAUSTEC praxis.

Laurini, R. (2000). A Short Introduction to TeleGeoProcessing and TeleGeoMonitoring. TeleGeo 2000, Sophia Antipolis, France.

Navratil, G. and A. U. Frank (2005). Influences Affecting Data Quality. ISSDQ, Peking, China.

Nothegger, C., S. Winter and M. Raubal (2004). "Selection of Salient Features for Route Directions." Spatial Cognition and Computation 4(2): 113-136.

Raubal, M. and S. Winter (2002). Enriching Wayfinding Instructions with Local Landmarks. Geographic Information Science. M. J. Egenhofer and D. M. Mark. Berlin, Springer. 2478: 243-259.

Retscher, G. and M. Thienelt (2004). "NAVIO-A Navigation Service for Pedestrians." Journal of Global Positioning Systems 3(1-2): 208-217.

Rümenapp, J. and Gutsche, J.-M. Verbreitung stationärer und mobiler Zugangsgeräte zum Internet in der Bevölkerung – Status Quo und Prognose bis 2015. ECTL Working Paper 15. Technische Universität Hamburg-Harburg.

Winter, S. (2001). Weighting the Path Continuation in Route Planning. 9th ACM International Symposium on Advances in Geographic Information Systems, Atlanta, GA, ACM Press.

Winter, S. (2002). "Modeling the Costs of Turns in Route Planning." GeoInformatica 6(4): 345-360.

Winter, S. (2003). Route Adaptive Selection of Salient Features. Spatial Information Theory. W. Kuhn, M. Worboys and S. Timpf. Berlin, Springer.

Winter, S., E. Pontikakis and M. Raubal (2001). "LBS for Real-Time Navigation - A Scenario." GeoInformatics 4: 6-9.

The Transition from Internet to Mobile Mapping

Michael P. Peterson

Department of Geography / Geology, University of Nebraska at Omaha, Omaha, USA
mpeterson@unomaha.edu

Abstract

Although the wireless Internet is a natural and inevitable progression from the wired Internet, the two developed much differently. The wired Internet was based on existing personal computer and workstation technology that had begun in the early 1980s. Over 1.3 billion computers were already in use when the World Wide Web was introduced a decade later. The browser software was free and no additional hardware was required – beside the use of a modem for connection purposes. The Web was quickly applied for all sorts of information delivery and its use expanded rapidly. In contrast, the development of the mobile Internet is hindered by a number of factors, not the least of which was the development of a wireless infrastructure. To fund its development, commercial companies devised new services that would generate revenue. One such revenue scheme was Location Based Services (LBS), a model for informing the mobile phone user where they are currently located and what services are available in the surrounding area. For a variety of reasons, LBS has not grown as quickly as had been predicted. While one might blame the small display, there are many reasons examined here why the wireless Internet and LBS have not expanded more rapidly.

1 Introduction

Within little more than a decade after the release of Mosaic, the first multimedia web browser, the distribution of maps through the Web – either static or interactive – had become firmly entrenched and drastically altered the way that people access maps. The Internet, now described as history's most powerful communication tool, has had a profound impact on map delivery. Within a few years in the mid-1990s, the distribution of maps through the Internet grew from almost zero to an estimated 200 million a day. Never in the history of cartography has there been such a dramatic shift in the way maps are delivered to map users. While Internet cartography is still in the process of development, millions of map users now turn to the Internet to access all types of geospatial information.

Concomitant with the growth in the use of the Internet has been the introduction of mobile telecommunication. In recent years, numerous telephone/computer handheld devices have been introduced with varying levels of computer processing and telecommunications capabilities. An attribute shared by all of these devices is a small screen and this constraint has been particularly limiting for the display of maps. For this and a variety of other reasons, mobile mapping has yet to be accepted as a viable technology – besides the larger navigation systems available for cars. Even here, there are reports that most of the features of car navigation systems are not used because the user interface is complicated. In addition, the maps are often obsolete and obtaining updated information can be very expensive.

In contrast to mobile systems, the development of the Internet was based on existing personal computer and workstation technology which had started to be used in the early 1980s. There were over 1.3 million computers connected to the Internet (Kikta, et. al. 2003, p. 10), including 225 million personal computers (Computer Industry Almanac 2005), when the World Wide Web was introduced through Mosaic in March 1993. The browser software was free and no additional hardware was required, beside the use of a modem to connect to the Internet. In addition, the Web was quickly used for all sorts of information delivery from news to email, and from commerce to promoting the Jihad. There were few limits to what information was available.

Not to be overlooked is the fact that the personal computers that were in use at that time were mainly designed for text processing. Page layout programs were in wide use by the late 1980s. By the early 1990s, graphic software had been introduced as well. Personal computers were essentially used to help produce output on paper, consisting of both text and graphics.

In fact, concerns were expressed during this time that these so-called paper-saving devices were actually contributing to more paper usage.

Mobile devices developed much differently and were either not connected to the Internet – as with PDAs – or were tied to expensive commercial mobile phone technology. In an attempt to further increase revenue, mobile phone companies introduced Location Based Services (LBS), a model for informing the mobile phone user where they are located and what services are available in the surrounding area. The complexity of the system, combined with privacy concerns slowed it adoption and growth. The integration of GPS devices within the mobile device to more accurately locate the mobile phone user did little to further the development of LBS.

While the growth of LBS had mired, the growth of mobile devices expanded at an exponential rate. The Computer Industry Almanac (2005) reported that the worldwide number of cellular subscribers surpassed 2 billion in 2005 — an increase from only 11 million in 1990 and 750 million in 2000. Figure 1 shows how mobile phone use varies by region. Europe leads the world in mobile phones per capita followed by the US and the Asia & the Pacific region. China is the clear leader in the number of cellular subscribers and reached nearly 400 million in 2005. No other country comes close. Russia has seen tremendous growth in the last few years and is projected to be in third place in absolute numbers by year-end 2005. Rapid expansion in India will see a future climb in the rankings to a possible #2 in 2010. Worldwide cellular subscribers are forecasted to reach 3.2B by the end of 2010, approaching one-half of the world's population.

While the delivery of spatial information with mobile devices has great potential, it has not as yet met the monetary expectations of telecommunication industry. A variety of reasons are examined here for why the development of the wireless Internet, and spatial information delivery in particular, has occurred differently than its wired counterpart. These differences will continue to affect how mobile mapping develops in the future. We begin by examining the differences in the development of these two forms of Internet access.

Table 1. Information Technology Usage Trends (Computer Ind. Almanac (2006)

Cellular Subscribers Per 1,000 People	1990	1995	2000	2005	2010
USA (#)	21.1	127	388	683	946
W. Europe (#)	9.1	60	634	930	1,008
Asia-Pacific (#)	0.4	7.1	71	230	379
Worldwide (#)	2.1	15.6	123	319	478

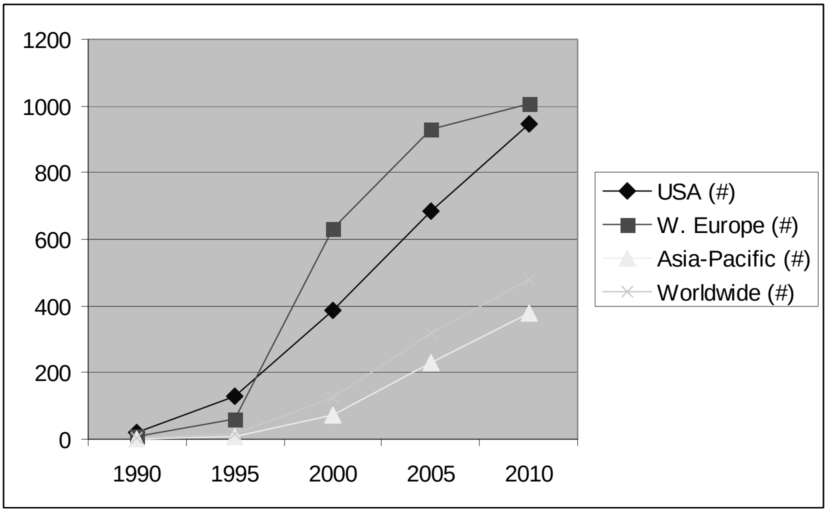


Fig. 1. Cellular subscribers per 1,000 people by region.

2 Contrasts in Development between the Wired and the Mobile Internet

2.1 The Role of Government

With considerable investment in infrastructure, the US government played an important role in the development of the Internet. The initial impetus derived from paranoia, a fear that a nuclear attack by the Soviet Union would render communications systems in the United States inoperable. In 1969, three computers, located remotely from each other, were connected under what was then Advanced Research Projects Agency Network (ARPANET). By 1971, ARPANET computers were connected at nearly two dozen sites, including Harvard University and MIT and protocols for remote terminal access (Telnet) and file transfer (FTP) were defined. As Kitka et. al. (2002, p. 8) point out, sometime during the 1970s ARPANET developed into a government-subsidized person-to-person communications service. In the 1980s, the Internet transitioned from the control of ARPANET to the US National Science Foundation. Established in 1986, NSFNET had a speed of 56 Kbps and 5000 Internet hosts. While Tim Berners-Lee posts the first computer code for the World Wide Web from Switzerland in 1991, many of the early developments of the Web during the 1990s also came from the United States. Foremost among these in the 1993 introduction of Mosaic, the first graphics-based Web browser. During the following year, traffic on the Internet expanded at a 341,634% annual growth rate. By 1996, when Microsoft Internet Explorer was introduced, there were 12.8 million Internet hosts and over 500,000 WWW Sites and the number of users reached 40 million. By 2001, the United States with less than 4.5% of the world's population represented over 40% of Internet users.

In contrast, the role of government, and the US government in particular, in the funding and development of mobile networks has been minimal. In Europe, governments provided funding and helped establish standards that would allow interoperability between systems. In the United States, a more laissez-faire approach led to many cellular systems with most being incompatible with the rest of the world. Duplicate cell phone towers were constructed by separate companies further increasing the cost of infrastructure for a country that is large in area. Kitka et. al. (2002, p. 95) state that the most important reason that the US is behind in mobile phone use is geographic size and population density. A smaller investment in infrastructure is needed in Japan and Europe because there is less space to cover and often more people in that smaller space. For whatever reason, the US is far behind other developed, and some lesser-developed countries. In 2002, it ranked 30th worldwide in cell phones per 100 population behind countries like the Czech Republic, Portugal, and Slovakia (see Table 1). Since 2002, a number of other countries have surpassed the US.

While the role of government in the development of technology is not always clear, most would agree that the wired Internet could not have developed in its present form without the funding provided by the US government. In addition, as the largest economy in the world, the US has played a key role in many technological developments in the past century – both through government initiatives and private sector competition. Its lack of direction in the development of the mobile Internet, and mobile phone use in general, has been particularly noticeable.

Ling (2004) comments on the explosive growth of mobile phone use and the interesting variation in national adoption patterns. Stuckey (2004) points out that much of the variation can be attributed to differences in the billing systems. In much of the rest of the world, the caller pays for the cost of the call while in the US it is more common for the cost of the call to be shared by the calling and called parties, putting the called party somewhat at the mercy of the calling and arguably diminishing the attraction of carrying a mobile phone.

Table 2. Cell phones per 100 population in 2002 (Source: NationalMaster.com)

Rank	Country	per 100 population	Rank	Country	per 100 population
1	Taiwan	106.45	23	Korea, South	67.95
2	Luxembourg	101.34	24	France	64.70
3	Hong Kong	92.98	25	Hungary	64.64
4	Italy	92.65	26	Australia	63.97
5	Iceland	90.28	27	Japan	62.11
6	Sweden	88.50	28	New Zealand	61.84
7	Czech Republic	84.88	29	Slovakia	54.36
8	Finland	84.50	30	United States	48.81
9	United Kingdom	84.49	31	Brunei	40.06
10	Norway	84.33	32	Canada	37.72
11	Greece	83.86	33	Poland	36.26
12	Denmark	83.33	34	Malaysia	34.88
13	Austria	82.85	35	Turkey	34.75
14	Spain	82.28	36	Thailand	26.04
15	Portugal	81.94	37	Mexico	25.45
16	Singapore	79.14	38	Philippines	17.77
17	Switzerland	78.75	39	China	16.09
19	Belgium	78.63	40	Indonesia	5.52
20	Ireland	75.53	41	Vietnam	2.34
21	Netherlands	72.24	42	Cambodia	1.66
22	Germany	71.67	43	Laos	1.00
			44	Burma	0.03

2.2 Content Development

If the Internet had no content, it wouldn't be of interest to anyone. Content makes the medium useful and interesting, and the easiest way to add content to a new medium is to copy the old. This fact was recognized as early as 1971 when Michael Hart (1992) started the Gutenberg Project for the online distribution of classical texts – the first being the U.S. Declaration of Independence. The idea was to convert the information on paper to the computer and thereby make it available to more people. The same thought occurred in the 1990s to map librarians at the Perry-Castaňeda Library Map Collection at the University of Texas as they converted their collection of maps to be accessible through the World Wide Web (University of Texas 2005). Most of the maps scanned by the University of Texas Libraries and served from their web site, currently

5715, were produced by the US government (CIA, USGS, National Park Service) and therefore are in the public domain. No permissions are needed for their distribution. The maps are available in ordinary JPEG format and hundreds of thousands are downloaded every day. Some are even available in PDF format that can take advantage of the better resolution of printers.

In contrast to the large amount of content available through the Internet, information available through mobile phones is very restricted. Among the many reasons for this are slower communications speeds and the need to re-format web pages for a smaller output size. In addition, the cost structure of mobile phone providers is such that access to web content is more expensive than ordinary phone service, either for the mobile phone itself or added charges that may be assessed according to the number of pages viewed. In short, the current mobile phone is not a good medium for information delivery for anything but the human voice, short text messages and small pictures. This situation will certainly change but it will require a considerable investment of capital and time.

2.3 Compatibility with Existing Media

Alan Kay (1977), who conceived of the Dynabook and whose design work led to the development of the graphical user interface, argues that the computer is not a tool or an instrument, but a medium. A medium is the carrier of information and is used to transmit knowledge and ideas between people. Each medium has a certain potential for communication. The computer, with the help of the Internet, is being used not only as a tool to help make maps, or search a database, but as a medium of communication. It was argued by McLuhan in the 1960s that we live in a rear-view mirror society (Theall 1971). According to McLuhan, all new forms of media take their initial content from what preceded them. Not only is the new medium based upon the old, but society dictates that the only acceptable way of approaching the new medium is by first emulating the old – through the rear-view mirror.

In order for a new medium to take its initial content from what precedes it, it must incorporate a delivery mechanism that can display that information. The personal computer made this possible in cartography for everything but large maps – the kind of maps that only a few had access to anyway. However, to preserve the property of mobility, wireless devices cannot display most of the content of paper or the traditional computer medium (desktop and portable computers) because of the small display size. Making the display larger would make the devices less than mobile.

Funk (2004, p. 44) estimates that doubling the display size doubles the display weight and increases the weight of the plastic housing by 50%. Similarly, a 20% increase in display area leads to a 30% increase in price. The issue, Funk goes on to point out, is whether customers will choose a heavier and more expensive phone in order to have a larger display.

As Rhoton (2002, p. 117) states, it would be convenient if all the applications that been developed over the past years on computers “could simply be dropped into a mobile environment and continue to work without any additional effort.” To be backward compatible, the mobile device would have to be able to perform the same function of the previous device in the process of adding new functionality. In addition to the small screen size, Rhoton (2002, p. 117) points out that mobile systems cannot be backward compatible because: 1) Many mobile platforms are closed systems, such as the Blackberry; and 2) there is a great deal of diversity in the machine interfaces of mobile devices. For these reasons and others, the mobile devices that are currently available represent a medium that is not backward compatible with either paper or the Web. Further, the medium may not adequately support the size and graphic requirements that have been traditionally associated with map use.

2.4 Developing a user base and paying for information

Rhoton (2002, p. 9) points out that a computer is not affordable for large segments of the world's population. Even those who can afford it may not feel the value of the Internet is high enough to justify the investment. Wireless devices are smaller and more affordable, at least initially, than desktop systems and more easily deployed in environments where space is limited. One of the reasons why the wireless Internet has been so popular in Japan is people’s mobility and the lack of extra space in people’s living areas.

Japan provides a good example for how mobile phones will impact our lives. In 2002, only about 15% of Japanese consumers and business people were using the Internet through PCs. The remaining 85 percent were willing to accept the limitations of smaller display screens and keyboards on wireless handheld devices (Kitka et.al, 2002, p. 101). It is estimated that the Japanese mobile Internet market represents more than 75% of the global market for mobile Internet services, with Korea a strong second (Funk 2004, p. 7). Mobile phones are already used in these countries as portable entertainment players, televisions, cameras, membership and loyalty cards, guidebooks, maps, tickets, watches, alarm clocks and devices for accessing everything from news to corporate databases (Funk

2004, p. 1). It is estimated that within a few years, mobile phones will be used in Japan for train and bus passes, credit and debit cards, keys, identification, and even money.

One reason for Japan's greater success is that the mobile telephone giant NTT DoCoMo created a separate organization to focus on consumers, as opposed to business users, and devised a micro payment system in which NTT DoCoMo collects charges for content providers (Funk 2004, p. 8). The company earns a 9% fee from providers that charge for their information (Kitka, et. al., 2002, p. 99). The lower emphasis placed on business users by Japanese and Korean firms made it easier for them to ignore business users and focus on general consumers in the mobile Internet (Funk 2004, p. 11). In the rest of the world, when it comes to new technology, business users are viewed as the early adopters and telecommunication companies cater only to their needs.

The user base for the Internet developed quickly during the 1990s but the remuneration of content providers on the web is still a major problem. There is still an expectation for free data and many providers of information have yet to devise a reasonable and workable payment system. This has stunted the development of online maps in that there is little reason to develop new content if cartographers will not be paid for their work.

2.5 Free maps

Paying for maps, for many people, has always been an option, not a necessity. In the US, state road maps are usually provided for free as a way to encourage tourism, and many tourist agencies provide local maps without cost. City maps are often available in the local phone book. Membership in an automobile club also provides “free” access to maps although the annual membership, including a variety of services, can be quite expensive. The Internet has further perpetuated the concept of free maps, with sites like MapQuest and Google making user-defined maps with only the cost of annoying ads. All of these free maps are based on street databases that were originally created by the US government and provided to the companies at no cost.

While base maps may be made available at a minimal cost in some countries, a considerable monetary investment is required to transform existing maps for use on small display devices. The maps need to be highly generalized and this is not a task that can be easily automated. Car navigation companies re-coup the cost of this new “cartography” through the associated display hardware. There is no special display counterpart with mobile phones and therefore no method to recover costs through the purchase of additional hardware. Providers will be forced to charge for the mobile map product, a practice that will be resisted by the mobile phone user.

2.6 Competition from wired Internet map use

The distribution of spatial information through wireless networks is in competition with more than a decade of development in Internet mapping. Three identifiable trends in this new form of map distribution have already emerged. The first era, from about 1993-1997, consisted of scanning paper maps and limited forms of interactive maps. The second era was dominated by both interactive street mapping (MapQuest along a series of competitors) and online GIS systems. The third development is community mapping, the ability of map users to change and update the content of online maps. Analogous to editing a website, as implemented through sites like Wikipedia (Wikipedia 2005), Google introduced a similar concept for maps allowing users to enter information (Google Maps 2005). Signs of a similar form of growth in the development of the wireless spatial information delivery are hard to find. As mobile mapping develops, it will have to contend with a more mature development platform in its wired counterpart.

3 Changing Patterns of Map Use

It is clear that new forms of map delivery will change how maps are used, and the resultant delivery mechanism.

3.1 Navigation

Many people consider navigation to be the largest potential market for mobile Internet services. An indicator for this is the popularity of car navigation systems. These systems are particularly popular in Japan where there are more than 10 million installed units, about 14% of all vehicles (Funk 2005, p. 125). But, these systems remain expensive. One reason for the high cost of car navigation systems, up to $4000, is that few of the components are standard. Most manufacturers use proprietary CPUs, maps, map engines, operating systems, and displays. Competition forces prices up by adding features such as improved displays, more detailed maps, and DVD movie playback.

Another trend emerging from Japan is that train, bus and destination-information services are far more successful than map services, the latter of which was expected to form the basis for a mobile Internet navigation service (Funk 2004, p. 132). Users of the destination-information services input departure and arrival times and the system responds with information from actual timetables. It is easier to download train information and information on restaurants in text form than to download map information. So, at least in Japan, destination-information and restaurant services have become successful while map services have not (Funk 2004, p. 132).

Mobile map services began in Japan in late 1999 (Funk 2004, p. 137). Similar to online maps with PCs, users input an address to generate a map of the site. The providers offer simple maps for free and charge for more detailed maps (about $2.50 a month). The most popular function is sending maps in mail messages by forwarding the URL of the map. Use of these services is far less than for the destination information services, probably on the order of only 1% in comparable traffic (Funk 2004, p. 137). Not only are the map services more expensive but the small screen and poor resolution make the maps hard to read.

An interesting development in mobile phone mapping in Japan involves the generalization of maps. The small display forces a heavier reliance on landmarks. But, even here, the map provider had to reduce the number of landmarks on the maps because of the available screen space. This reduction in the number of landmarks was too difficult to do manually so the companies relied on an automated system to decide which landmarks could be shown at certain scales. This caused well known landmarks in some areas to be eliminated. Although improved displays will allow more landmarks to be shown, the map provider will have to update the underlying database thus incurring further costs (Funk 2004, p. 137).

3.2 GPS integration

It is anticipated that the combination of mobile phones and global positioning systems will be the next major development in mobile mapping. However, the power required by the GPS unit is still a major problem. In vehicle navigation systems, there is sufficient power from the car to power the receiver. The power consumption of the GPS receiver makes the use of pure GPS with phones more difficult and thus requires some form of "assisted GPS." In A-GPS, most of the GPS calculations are done at the server level. The receiver, being limited in processing power and normally under less than ideal conditions for position fixing, communicates with the “assistance server” that has higher processing power and access to a reference network of locations. The main purpose of AGPS is to provide municipalities with location-based emergency phone service, such as emergency 911 service and designed for short-term use (Wikipedia 2005). A-GPS reduces power consumption and increases positional accuracy but drives up communication costs. In some markets, current communication charges would mean that if a phone's location were updated every 5 minutes over an 8 hour period, a charge of over $80 would be incurred. A-GPS position may also time-consuming. It takes approximately 15 seconds to download the GPS data, 15 seconds to send these data to the server and carry out the calculations, and 15 seconds to download the map to the mobile unit (Funk 2004, p. 145). Most users would not want to wait 45 seconds.

Put into effect at the end of 2005, the US Federal Communications Commission (FCC) now requires all cell phone carriers to provide the ability to trace cell phone calls to a location within 100 meters or less. In order to comply with these FCC requirements, most carriers have decided to integrate GPS technology into cell phone handsets, rather than overhaul the tower network that could be used to triangulate the position. Most GPS enabled cell phones do not allow the user direct access to the GPS data. Rather, the location data will only be sent to the cell phone provider if an emergency call is made. As the system is being designed for emergencies, it seems unlikely that such a system will provide continuous GPS information – a requirement for effective mobile mapping.

3.3 People Tracking

People tracking involves the use of a GPS device to track the location and movements of people. Tracking the location of children, or parolees, or even pets, may provide the impetus to provide the necessary GPS and online tracking infrastructure. Several service providers advertise the capability to track a person or an object on a PC screen. Indeed, most commercial trucks are tracked in such a way. Alan Philips of uLocate.com is optimistic about the future of people tracking, saying “Cell phones are becoming multi-purpose devices capable of much more than simple voice communication. By leveraging the latest in satellite and wireless technology, uLocate is providing a secure, reliable, service that creates value for families, individuals, and small businesses without any additional overhead, hardware or software” (Directions 2003).

Monmonier (2003) examines how mapping is being used to invade privacy. He points out that: “Web cartography is especially valuable—and potentially threatening—because it not only greatly expands the audience of potential watchers (Peterson, 2000) but also allows for unprecedented customization of maps that describe local crime patterns, warn of traffic congestion and inclement weather, disclose housing values, or—thanks to the Global Positioning System (GPS) and the new marketplace for “location-based services“—track wayward pets, aging parents, errant teenagers, or unreliable employees” (p. 98). He concludes that: “As society and government work through the significance of locational privacy and decide what legal limitations, if any, are appropriate and permissible, the debate will turn to possible restrictions on Internet cartography…” (p. 111). While the tracking of people may become technically possible, it will raise concerns among privacy advocates. User-centered mobile mapping may not be allowed based on these concerns.

4 Mental maps and wayfinding

A common misconception is that map use occurs at the instant that the user is examining the map. To think this way is to deny that humans have a memory and are able to process information after an object has been removed from view. It can be argued that the majority of map use occurs without a map being present. Our daily movements within our surroundings, indeed, any thought about the spatial world beyond our immediate view, is based on both first-hand experience and the maps we have seen. Although the depictions of the spatial world we keep in our brains are not always accurate, they suffice to help us find our way through our immediate environment and to think about the world beyond. These *mental or cognitive maps* are the *maps* that we actually use. They are also the representations that give us the comfortable feeling that we know where we are and how to get to places. Mental maps help us feel “found” in the sense that we are aware of our spatial surroundings. In contemplating the future of Internet maps and mobile mapping, we need to examine how maps influence the internalized representations we form from them, and how they help us interact with the world.

4.1 Mental map formation

While we don’t know exactly how mental maps form, we do know that they are based both on direct experience with the environment and looking at representations of the environment in the form of maps. The internalization of maps seems to happen without conscious awareness because people can recall the outlines of countries and the shapes of continents without intentionally remembering to do so. In addition, these shapes are remembered quite early as children can recognize outlines of countries and continents at young ages. Maps have a major impact on the formation of mental maps from a very early age.

4.2 Mental map formation in wayfinding

There seems to be close relationship between the formation of mental maps and the process of wayfinding. While the primary purpose of wayfinding is to get to the destination with as little effort as possible, the secondary purpose is to create a mental map of the route, however primitive as first, that will aid in finding the location again without the use of a map. In other words, the purpose of the map in this function is to create a mental construct such that it will be rendered meaningless when the same task is performed again. The map succeeds by becoming useless over time.

4.3 Creating a Permanent Dependence

In contrast, when using a mobile device for wayfinding, the user is directed to a location with minimal mental effort by the user. Because there is little coordination between the map and the environment, a functional mental map is not formed. In addition, the depictions presented on the mobile device are too schematic to create a useful mental map of the environment. It is very likely that the user will need to get instructions from the device again for not only the return trip but any future trip to the same location. What the mobile device has succeeded in doing is creating a permanent dependence on the device. In contrast to a map, the mobile device succeeds by becoming indispensable.

Being told where you are bypasses the process of finding out where you are and this inhibits the formation of mental maps. Certainly, the process of finding your location on a map helps to form a mental map, a mental conception of where you have been and where you need to go. We need to be careful in developing wayfinding tools that we also contribute to the formation of long-term mental maps.

5 Conclusion

There are major differences in the development of the wired and wireless Internet and these differences will affect how maps and spatial information are delivered to users in the future. The wired Internet was an outgrowth of a government initiative while the wireless Internet is almost totally a commercial enterprise. The commercial aspect of the wireless Internet permeates all aspects of its growth and development.

As Kitka et. al. (2005, p. 106) state, the “wireless Internet will most certainly provide society with new experiences and freedom, and unprecedented access to information.” Funk (2004, p. 51) sees that larger “memory, faster processing capability, and greater network speeds are being used to improve the user interface, and it is possible that they can be used to significantly improve the user interface in the future.” Wiberg (2005, p. 344) adds that the “computers of yesterday were about getting things done as quickly as possible, the computer of tomorrow will help us to prolong, sustain and develop the things we care about the most, i.e., our ongoing interactions with others.

The size of the display remains a concern for the delivery of maps. The most troubling aspect of our dependence on this new technology in cartography is that it may not provide long-lasting spatial information. Rather, the systems will help guide us to a location without providing the type of long-term information that will help us find that location on our own. Further, this dependence will lead to an uneasy relationship with the world around us – and perhaps the feeling of always being lost. If we are to best use these new technologies, we need to remember that the ultimate purpose of a map is to provide a sense of connection with the world around us.

References

Computer Industry Almanac (2005) China Tops Cellular Subscriber Top 15 Ranking. http://www.c-i-a.com/pr0905.htm

Computer Industry Almanac (2006) Europe #1 in Per Capita Cell Phone Usage. http://www.c-i-a.com/ pr0206.htm

Directions Magazine (2003). New Service Enables Parents To Pinpoint Kids’ Whereabouts Using Cell Phones. [http://www.directionsmag.com]

Funk, Jeffrey L. (2004) Mobile Disruption. Hoboken, New Jersey: John Wiley.

Google Maps (2005) [http://maps.google.com/]

Hart, Michael S. (1992) *History and Philosophy of Gutenberg Project.* [http://www.gutenberg.org/about/history]

Kay, A., and Goldberg, A. (1977) Personal Dynamic Media. Computer (USA): 31–41.

Kikta, Roman, Al Fisher & Michael Courtney (2002) Wireless Internet Crash Course. New York: McGraw-Hill.

Ling, Richard (2004) The Mobile Connection: The Cell Phone's Impact on Society. San Francisco: Morgan Freeman Publications.

Monmonier, Mark (2003) The Internet, Cartographic Surveillance, and Locational Privacy. In Peterson, Michael (ed.) *Maps and the Internet.* Amsterdam, Cambridge: Elsevier Press.

Nationmaster.com (2002) *Mobile Cellular per 100 population.* [http://www.nationmaster.com/graph-T/med_mob_pho&int=-1]

Rhoton, John (2002) The Wireless Internet Explained. Boston: Digital Press.

Stuckey, John (2004) Review of The Mobile Connection: The Cell Phone's Impact on Society. *Ubiquity*, Volume 5, Issue 28, Sept. 8 - 14, 2004. http://www.acm.org/ubiquity/

Theall, Donald F. (1971) The Medium Is the Rear View Mirror: Understanding McLuhan*.* Montreal: McGill-Queens University Press, 1971.

University of Texas (2005) Perry-Castaňeda Map Library. [http://www.lib.utexas.edu/maps/]

Wikipedia (2005) Wikipedia, the Free Encylopedia. [http://en.wikipedia.org/wiki/Main_Page]

Theory and development of research in ubiquitous mapping

Takashi Morita

Department of Civil and Environmental Engineering
Hosei University, Tokyo, Japan
morita@k.hosei.ac.jp

Abstract

Understanding of the emerging field of Ubiquitous Mapping requires a clear definition of the field and the list of issues to study. The author proposes a simple principal framework of the field, and provides a discussion of the future direction of research. Ubiquitous Mapping refers to the ability or the environment within which users can create and use maps anywhere and at any time to resolve spatial problems. The most characteristic element of this approach is that of context awareness. Geographic Information Systems (GIS) tools differ from those of Ubiquitous Mapping in that GIS is oriented toward spatial analysis in a defined space whereas Ubiquitous Mapping is concerned with spatial communication and is comprised of the four basic elements, the real world, the map, the user, and the information technology (IT) infrastructure, in a human-oriented context that emphasizes the interaction between the output map and human responses such as spatial cognition and decision making. Essential to this process is context awareness. Thus, in Ubiquitous Mapping the notion of a “map” must be replaced with that of “mapping” since the user demand is for the real-time creation and manipulation of maps. Consequently, “maps” must therefore include not only 2D, but also 3D views, such as those where users observe the real world from the side in order to recognize and verify the various spatial elements of a particular

scene. Also, because 3D maps require common spatial articulation, the abstraction and categorization of objects, as well as a clear reference system relating the features of the map to real objects is often difficult, particularly in urban areas where many objects have similar shapes. This phenomenon is conducive to the development of the notion of the real scale map. A further difference is that traditional maps only describe immobile features, while Ubiquitous Mapping can be used to represent human activities such as the planned route of travelers and other events may be represented in real time.

1 Map Communication

Ubiquitous Mapping refers to the use and creation of maps by users anywhere and at any time. It is strongly influenced by advances in information technology, such as the development of wireless systems, high-density data storage and broadband communication, which can be seen as stimulation and facilitation of dynamic and personalized mapping. However, it appears that the fundamental nature of the map has not been changed significantly from its origin. Before examining the notion of Ubiquitous Mapping, it is necessary to review the purposes for which maps have been designed.

1.1 Three elements

At a fundamental level, maps can be considered as providing a framework for depicting location. This location may be relative, a relationship between known and unknown elements, an absolute location, or a coordinate system. This is the most fundamental characteristic of a map. The second characteristic is that maps are primarily represented in visual format. We can easily recognize the relationship between map elements because they are reconstituted in our brain as an image with a spatial component. The ease with which patterns can be recognized is another characteristic of a map. However, the accuracy required to represent spatial data correctly has resulted in the need for protocol in cartography. The third characteristic is the human-map-space interaction, which can be traced back to early human history.

1.2 Origin of Map Communication

These elements are depicted in Figure 1 where one person is describing the location of an object to the person facing to him using a map on the ground. The object is located beside a route just before the road crosses a river (relative location). It is represented visually by employing different lines, of which each represents a structural pattern. Communication of an object's location is easier and quicker using visual means compared to verbal explanations. Given that this form of communication can exist independently of technological developments, it is likely that communication activities among humans have always employed these techniques.

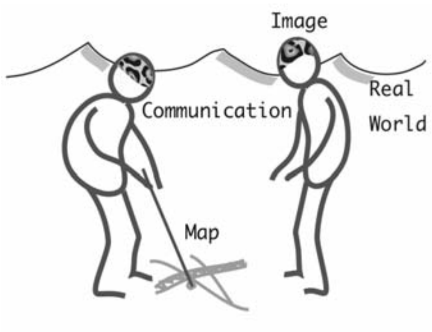


Fig. 1. Origin of Map Communication

There are three basic components in the aforementioned situation: a map, spatial images in the brain, and the real world.

1.3 Scheme of Map Communication

Figure 2 is a diagram depicting the notion of map-based communication proposed by Lech Ratajski [1] and adapted by the author [2]. For map-based communication, there is a sender and a recipient of information. When the sender wants to send map information to a recipient, then that sender has a spatial issue to resolve and one of the ways in which this can be done is to use a map. Once a subject is fixed, the sender collects the information necessary to build a map. The information comes from various materials, databases, and even from other people when seeking assistance. Once sufficient information has been collected and collated, the next step is the design of map symbols and appending a legend. This is very important because the effectiveness of a map depends largely on the design of symbols.

Understanding the contents of the map by the recipient begins with the perception of a spatial image forming in the brain after viewing the map. This is a cognition process using the distribution pattern of map symbols.

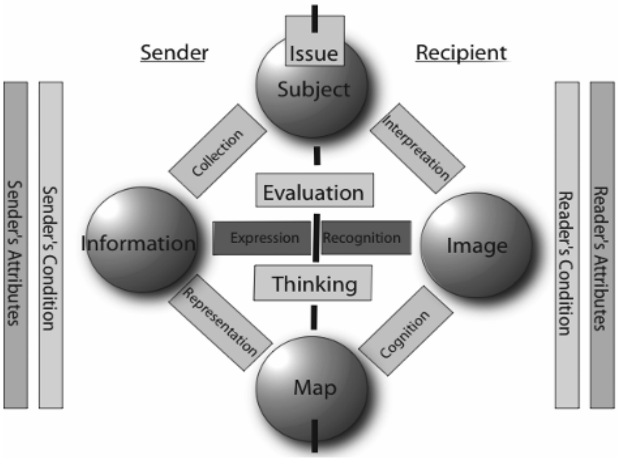


Fig. 2. Map Communication

As shown in figure 2, a recipient begins to extract the geographic meaning not only at a detail level (e.g., location of an object through map symbols), but also at the level of the entire image (e.g., I am here on the map, how can I get to the object?). In the latter step, the recipient interprets a map relatively to his extant situation; the context. The sender wanted to convey the location of an object while the recipient sought to determine the location of the object and also how to get there. In this case, representation of an object's location, which may be a subsidiary function, has been achieved. However, if this function is not satisfying, then it is a bad map that renders map-based communication impossible. Map-based communication between a sender and a recipient is considered effectively, if the object the sender intended to show was recognized by the recipient. Generally, absolute success may not be expected where a raison d'etre for a cartographic study exists.

2 Mapping World

Paper maps can be carried in the field. The information on the map is "fixed" on the paper and it is the user who must extract the required information from the map. However, it is not always possible to find the necessary information because there is a limit to the surface area that can be dedicated to depict the information on the map. There are also differences between maps and reality because of the limits imposed by the frequency of maintenance of updating information. Furthermore, making map production viable demands that a large number of copies is printed within a certain time frame. Thus, there is a need for the development of a new map, one that can be produced on site. Moreover, if it is interactive and can satisfy the specific needs of the individual who requires the map, production can be referred to as on site “mapping”.

2.1 From “Map” to “Mapping”

Figure 3 shows the mapping world [3] with different types of maps depicted in the cube. This model is composed of three axes that are defined by C. Morris as the semantic, syntactic and pragmatic dimensions [4]. In the model, the semantic axis represents spatial articulation (standard vs. individual), the syntactic axis corresponds to the spatial representation (multi-dimensional vs. two-dimensional) while the pragmatic axis represents the type of media (fixed vs. flexible). A paper topographic map would be located in the far lower corner of the cube given that it is a two-dimensional spatial representation on fixed media with standard spatial articulation. In the uppermost corner of the cube is a multi-dimensional spatial representation on flexible media with individual spatial articulation, representing the maps of the future. Between these end-points lie systems such as GIS and car-navigation. All map types can be classified within this mapping space, either as a single mapping system or as a set of different systems with different objectives.

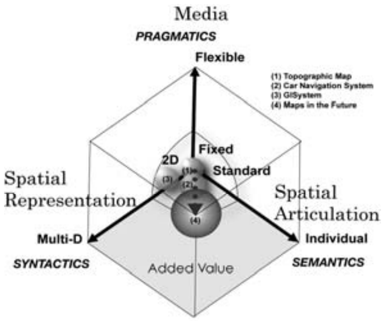


Fig. 3. Mapping World

Placement of the topographic map in the corner opposite future maps does not mean that the topographic map format is an "old" map because topographic maps already exist in digital format. Rather it means that topographical maps represent the standardized basic map format relative to all other types of maps. Without a basic map format, there would be no common space to improve upon and value-added, systematized maps would not have become possible. Furthermore, the re-evaluation of information from old maps, such as hand-drawn and externalized cognitive maps, is an ongoing process. If a standard base map exists, old maps considered to be inaccurate may be relocated onto it. The base map then becomes a platform upon which various mapping information can be manipulated using modern information technology (IT) systems. However, these inaccurate maps could have significant value as human interfaces. Use of a well-drawn albeit out-of-perspective map - a good generalization in other words - can often meet with favorable results. On road maps for example, the characteristics of the road form are often simplified or exaggerated to make the map easy to read and understand. This illustrates that homogeneous and accurate maps are not always efficient for human interpretation.

2.2 Egocentric mapping

The development of mobile phones equipped with GPS and digital compasses, as well as car-navigation systems, can be considered progenitors of future mapping methods. These mapping methods are personalized, bi-directional and change in real time. The most important characteristic of this style of mapping is that it is egocentric, meaning that a map may be presented on demand in relation to the actual position of the user. “Where am I?” is always the first question any user asks when using a map on site. If the direction of the map is always adjusted to the north, then there is parallelism between the map and the surrounding landscape and if a map shows a side view of the users’ actual position, it is easier for the users to orientate themselves. Individuals differ in the way they perform spatial tasks, with actions such as going to the bookstore, restaurant, or flower shop, all depending on the situation. This is the notion of context awareness, and it is likely to a key term in future mapping.

3 Fundamentals of ubiquitous mapping

3.1 Changes in the information and communication environment

Effective communication is facilitated by a physical medium to convey messages between a sender and a recipient. The images in the brain should be externalized and tangible. Drawing a map on the ground with a stick gives the map a real shape. The combination of the ground surface and a stick can be viewed as the technical environment used to facilitate the representation of a map, one that may either become redundant or change as new techniques become available.

3.1.1 Visualization of information

Tribal humans were using maps to communicate spatial concepts considerably before the invention of letters and characters. However, after the development of typesetting it became easier to communicate by using text than by using graphics. This development resulted in printed characters dominating books for more than five hundreds years. However, subsequent to the development of printing graphics, books came to contain more than only text. Similarly, broadcast communication began with radio using spoken messages before evolving into TV, which incorporated the visual component in addition to sound. It is currently the same for cellular phones, personalized computers, and PDAs; portable devices that were previously incapable of sending visual information are now used to sending image-rich content. As can be inferred from this convergence, communication with words and visual information are both very important, with the latter being more difficult to represent.

3.1.2 Ubiquitous Computing

Computers and communication networks are similar to extensions of our cerebral and nervous systems, we have come to expect to use them anywhere and at any time. This situation reflects the goal of developments in information technology. Mobile devices such as cellular phones and PDAs have become pervasive, telecommunication lines are no longer made of copper wire, but rather are made of glass optical fiber for broadband - and now even wireless. Computers are ubiquitous and can be found in Internet cafés, airports, train stations, hotels, homes and offices, and these are just traditional applications. We already utilize ATMs (Automated Teller Machines) at banks, ticket-vending machines, route-guiding machines, in car navigation systems, and other machines that incorporate computers for specific purposes. Small IC tags that can be attached to fixed objects such as signposts, or moving objects such as consumer goods, may become information and communication stations, connecting different computers. These are examples of the so-called "ubiquitous computing environment". In this environment, communication between person-to-person, person-to-machine, machine-to-machine (machine communication: network, human communication: understanding) is bi-modal, interactive, and realized in real time. This is more than an information system, it is a communication system.

3.2 Ubiquitous nature of maps

The contemporary needs for ubiquitous mapping are mentioned above. Now we will examine the potential applications of ubiquitous mapping and its principal characteristics.

3.2.1 Visual perception

Maps are generally visually perceived. Visual perception is ubiquitous because we can perceive and conceptualize an entire object through pattern recognition. At the same time, however, we can observe any given part of the image. This means that we can move freely between the entire image and the partial image. The order of processing data is not predetermined as it is when we are confronted with reading text composed of characters. This ubiquitous nature of vision is very useful for reading maps as thematic images can be superimposed in the background (figure and background) on the same plane. It is thus possible to represent alternative propositions or solutions using the same image, facilitating the rapid and simplified communication of spatial information.

3.2.2 Creation process

We continually refresh the spatial and temporal attributes of images in the brain and modify these images as necessary depending upon the prevailing situation. Such maps are regenerated from information contained in a database to produce one of two kinds of maps: one is the reproduction of existing map, the other is the creation of a new map in response to specific user requirements that differ in time, space, and personal attributes. This latter map-type can be created using digital systems with specific parameters applied to existing maps that are modified and to which necessary information can be added. Realization of map production in this way has only become possible through enormous advances in information technology.

3.2.3 Using process

A map drawn on the ground surface is a type of on-demand map. A paper map can be transported and utilized by the user at a different location. Such a map can be used any place else including its place of origin. However, in order to solve a spatial problem successfully one should attempt to acquire a suitable map. This is usually done by specialists and is difficult for members of the general public. Consequently, it is therefore highly likely that the full potential of maps has not been realized. We have mobile and wireless devices that can display maps and can receive existing maps on demand, and we can also create maps in response to the needs of a user in a particular situation. This enforcement-assistance relationship between map and user was not possible previously. Maps too have become ubiquitous and can be made on-demand.

3.3 Difference between GIS and Ubiquitous Mapping

How does ubiquitous mapping differ from GIS or simple geographic information? In the mapping universe these concepts fall within the same schema. However, if we attempt to distinguish between these systems, it seems that the primary difference is related to the context of “information” and “communication”. As shown in Figure 4, GIS emphasizes information processing through data input, database building, data analysis, and data output of spatial information. It is used for spatial analysis using database and modeling toolbox type applications with maps simply being used as the outputs of this information processing.

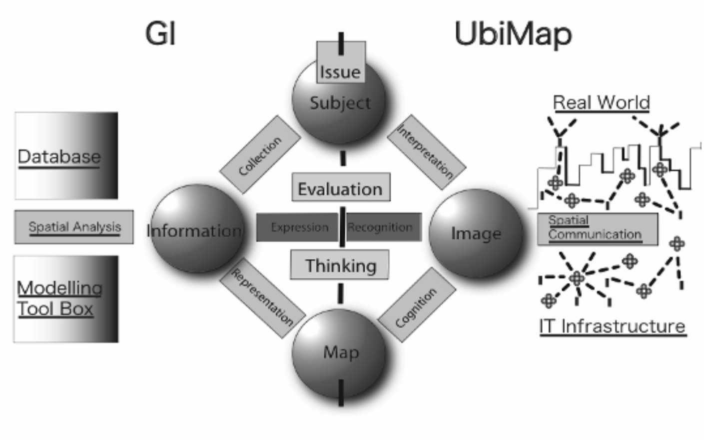


Fig. 4. GI and Ubiquitous Mapping

Conversely, communication is the principal function of ubiquitous mapping. It includes not only map production, but also map use and map communication and considers the interaction between the map, the spatial image, and the real world. Instead of being a "modeling toolbox" as in GIS, there should be "IT infrastructure" for ubiquitous mapping. It is the on-site communication network system connecting the three elements; the map, the spatial image, and the real world. Thus, GIS is oriented toward being a spatial information system for analysis, whereas ubiquitous mapping is concerned with spatial communication. Ubiquitous cartography is a human-oriented system that incorporates spatial cognition, spatial deduction and abduction, and spatial decision making.

4 Research in ubiquitous mapping

What is the future direction of this domain for further studies? Since the domain is in its infancy, we have attempted to provide a working framework and provisional research agenda below. Topics such as social costs, security and cross cultural studies will also be required in the future.

4.1 Elements of context awareness

Since the purpose of Ubiquitous Mapping is to resolve spatial problems and to realize the ability, and consider the environment of users in the creation and use of maps at any time or location, the most important and novel theme in cartography is the element of context awareness. This element is fundamental because real time mapping applications that are sensitive to context and, "anywhere, any time, and anybody" require that the components of the "where, when, and who" of the users is known in advance. The system should therefore recognize the following elements.

4.1.1 Three situations

at anywhere → where
at anytime → when
for anybody → who

Fig. 5. Where-when-who

Anywhere: To respond to the question of where you are, one may react using either natural language to indicate the relative position to a known landmark, place name, address, or even a geographical coordinate in an extreme case, or point to a corresponding site on a map.

Anytime: The users' temporal characteristics are usually "now", in real time and immediate. However, one may refer to a chronological point in time that differs in a relative sense from the time at which the question was asked, or to a difference in the standard and/or calendar time. Time may be defined using natural language such as in an emergency, the daytime or at night, for example.

Anybody: This includes the attributes of the user and corresponds directly to the objective of the mapping exercise. Characteristics such as gender, age, whether the user is a national of the area or a foreigner, whether they have a mental or physical disability, people in a state of haste and similar characteristics, as well as their intended activity, prescribe the function of map.

4.1.2 Four elements

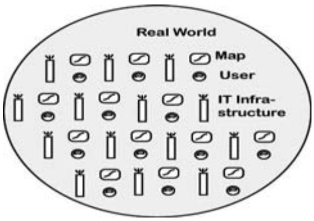


Fig. 6. Four Basic Elements of Ubiquitous Mapping

Ubiquitous Mapping consists of four basic elements: the real world, the map, the user, and the IT infrastructure with interactions between each element.

The real world: Is taken as the total space required by human beings for living; it is the space consisting of the objects and the backgrounds of our activities. If the characteristics of this space are different, human behavior would also be different.

The map: It is a physical map that is observable by a human sensor and is represented using map symbols. Voice may also be used to augment the data contained in the map. It may be provided on site as a guide map post or on the screen of a mobile device. Here, the notion of a "map" may be replaced with that of "mapping" as digital maps and IT infrastructure allow for real-time creation and manipulation of maps.

The user: It is the subject and the actor who needs to resolve a spatial issue. Individuals have spatial images in their brains that differ from those of other individuals. This is because personal attributes such as objectives, interpretations of spatio-temporal phenomena, personal history and spatial preference, all differ from person to person and are dynamic depending upon a given context.

The IT infrastructure: This is the supporting, spatial communication system that facilitates the interface between the real world, the map, and the user. It is composed of a wireless communication network, spatial communication devices, spatial information databases, mapping software, and similar attributes. Spatial communication devices can be divided into mobile devices and communication posts. The latter are stationary and on-site, and connect the real world, the map, and the users like a street name and house number are used as coordinates on a map.

4.1.3 Side view (3D view)

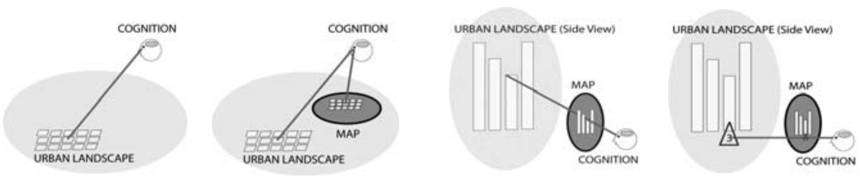


Fig. 7. Map and Spatial Cognition

The relative position between four elements mentioned above is also central to the context and usability of a system. When people view a townscape, the scenes are usually perceived as side or oblique views, very rarely they are viewed from a vertical angle. Since maps are normally drawn using a vertical projection with the element of cartographic abstraction, users are forced to interpret the side of a location into vertical view, a process that can demand considerable gymnastics of the brain. Thus, it is more convenient for a user, if the map is first drawn from a side view perspective and then transformed into a vertical view. However, the characteristics of side views vary with respect to the relationship between the viewpoint and the view angle of the object scene. Consequently, it is not possible to make a side view map from vertically derived cartographic materials. However, the considerable improvements in digital cartography have resolved many of the difficulties formally associated with this problem. We now have 3D urban models with which it has become possible to transform images from side views to vertical views. We have thus gradually resolved many of the problems formally associated with the cartographic abstraction required for generating side views. Photo-realistic images may not be the solution because such methods are always open to misidentification by the user. The object that constitutes “what” is a part or an element of a continuous space. A clear image of spatial elements may be given by cartographic abstraction, which may then be superimposed on to a photo or real scene. This latter approach is a sort of mixed reality where symbols like arrows are directly displayed on the site or on the mobile screen.

Fig. 8. Mixed Reality

4.1.4 Time representation (4D view)

Time and space are intricately connected, particularly in the context awareness environment because “when” is one of three fundamental parameters referred to above. Three time variables need to be represented. One is that of “real time” where the map on screen is refreshed relatively to the location of the user in real time. The second type includes the representation of mobile objects or temporal phenomena that are normally omitted from traditional paper maps. Location monitoring of children using GPS and a wireless network is one such example. Temporal information can be depicted on map-enabled cellular phones and transferred to friends to notify them of a destination to which they should come. The third temporal aspect is an animation showing the movement of a phenomenon. Animated instructions help users to make a decision, such as route guidance using dynamic arrows.

4.1.5 Real scale map

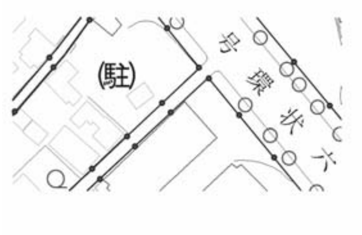


Fig. 9. Reference points

These maps have a scale that is unusual because it is 1:1. The abstraction of spatial objects and reduction of the size of spatial phenomena are some of the basic functions of map. However, abstraction really only begins at scales of spatial cognition when we can observe an object scene and recognize the spatial characteristics of the structure. This is so called spatial articulation and is performed by everyone in their daily lives, which is the level at which spatial elements can be distinguished and classified. This process is very similar to cartographic abstraction but it is done using real scale objects and with images being retained within the brain. A map is a product of cartographic abstraction of the real world. When we utilize a map, the on site object depicted by a map must be referenced to real object to verify a position. However, such referencing is not always easy to reference abstractions with the real world using only the characteristics of the feature being represented. Consequently, we often employ signs that are located on site as well as those, like street names, which can be described on a map. However, because this approach is not always successful, especially at local scales, it is necessary to have a more efficient system.

The real scale map proposed here is a map drawn on the real world using well-defined reference points that are visible on site and, simultaneously, as points on a map. Once this has been achieved, all cartographic elements should be defined relatively these reference points.

The difference between this system and the coordinate system is that the reference points on site are either visible or invisible. Consequently, reference points should be clearly visible and easy to locate. In addition, the orientation of the map should be always synchronized to the north and this could be supported by IT infrastructure.

4.1.6 Connection and transformation between different spatial languages

Location may be represented using natural language, such as address, place name or as “ahead”, “turn right”. Alternately, they may be also indicated on a map, which is more conventionally used to represent spatial information. Thus, while natural language is easier to present, it is less precise than cartographic language. There is thus always a need to transform spatial information between verbal and non-verbal formats and the process of geocoding constitutes the connection between these two spatial languages. In the context awareness environment, both natural and cartographic languages need to be understood as the system should be able to approximate certain parameters on behalf of the user in a given context and also so that users can use both languages unconsciously.

4.2 Applications in Tokyo

4.2.1 In-car navigation systems

A commercial system was introduced in 1981 without any digital information infrastructure. It functioned using an autonomic system that employed a dead reckoning method with gyrocompass and a paper road map. In the mid 1980’s, a beacon system that transmitted traffic information from dedicated transmitters along roads combined with digital road network data for whole country constituted the infrastructure. The use of GPS in 1990 combined with a map-matching method, as well as information on traffic congestion and accidents in real time using FM, subsequently became available. In 2005, more than 15 million in-car navigation units were sold in Japan. Their typical functions include the display of actual geographic location on digital maps in 2D and 3D with landmarks using north up/heading up, congestion/construction information in real time, route guidance (maps, diagrams and voice) to the destination using customized user profiles, queries and different display types (gas station, parking, restaurant, etc.) are also possible. The most recent systems use a hard disk, which calculates appropriate route to a given destination, even when the route is suddenly altered from the initial course.

4.2.2 Cellular phones

“All-in-one” type high-functionality cellular phones have become more popular than PDAs (Palm and Pocket PCs) in Japan. More than 90 million units have been sold and the diffusion rate is more than 70%. Typical specifications of current cellular phones include full-color LCDs, high-resolution display (240 x 320 pixels), 3D graphics engine, Java/flash/SVG compatibility, camera (3M pixels), removable memory, mail, Web, internet, voice recorder, a diary, 2D Bar (QR) Code Reader, GPS, Compass, etc., which can be applied as a human navigation system for route guidance by pedestrians. It proposes alternative routes after considering the various possible modes of transport (on foot, by taxi, by bus, by train, etc.) and provides timetable and tariff information. Once the route has been fixed, the user is provided with step-by-step directions and guided by voice with the map in the heading-up orientation. The destination may be designated directly using an address or a point on a map, or through querying the system using several destination categories. One can communicate one’s actual position using the GPS and a map to the recipient without knowledge of the exact address. Another location-based service involves transmitting the location of the bearer of the equipment (children, grandparents, etc.) to the user via an operation center. The equipment is a modified version of a cellular phone that transmits location information in response to demands sent from the operation center. The user can use the system to determine, if the bearer is within a predefined safety zone or the user can request the bearer to push a button to confirm that the bearer is not in trouble. If there is no reaction, a member of staff or security services will rush to the site.

Fig. 10. Human Navigation

4.2.3 IC Tags for navigation

The free Mobility Experiment Project (http://www.tokyo-ubinavi.jp/) organized by the Tokyo Metropolitan Government, the Ministry of Land, Infrastructure and Transport, the Japan Institute of Construction Engineering and the YRP Ubiquitous Networking Laboratory is undertaking a public experiment in Tokyo and Kobe. The objective of the experiment is to verify the functionality of equipment designed for ubiquitous computing that is being applied to navigation in a real city space. The system consists of an IC tag, wireless marker, infrared marker, ubiquitous communicator, and an electric cart to transport visitors, that all provide route guidance and information necessary for sightseeing. The ubiquitous communicator is a modified PDA with a GPS receiver, wireless antenna with preloaded information detailing the information relevant to the site of interest, identified when the PDA communicates with the IC tag (non-contact and passive) and a wireless marker (10 m active zone) at the site. The relationship between ubiquitous communicator, the IC tag and wireless marker constitutes a location-based context where the real space meets with mapping space. If IC tags and wireless markers constitute a map in real space, then this relationship using various location information becomes a real scale map. A reduced map of this map with the same reference points is thus generated and a user can refer to two points on the map easily and correctly. This relationship is conducive to the creation of mixed reality.

Fig. 11. IC tags located in front of a temple

5 Conclusion

In conclusion, ubiquitous mapping is more than a digital map distribution system that aims to provide technical solutions associated with map creation and use. Ubiquitous mapping accelerates, facilitates, and stimulates the universal nature of map creation and use through the application of advanced information technologies. Consequently, despite being derived from the most primitive map-types, ubiquitous mapping can be considered as representative of the future map type. Further studies should investigate the potential of applying new technologies and increasing our understanding of the nature of maps.

References

[1] Ratajski, L., 1973, "The Research Structure of Theoretical Cartography, International Yearbook of Cartography", Vol. XIII, p. 219

[2] Morita, T., 2002, "Reflections on the Results of a Questionnaire on the Definitions of Map, Cartography and GIS", The Selected Problems of Theoretical Cartography 2002, Proceedings of a seminar of the commission on Theoretical Cartography, Gdansk, Poland, July, pp. 34-44

[3] Morita, T., 2004, "Ubiquitous Mapping in Tokyo", Proceedings of the First International Joint Workshop on Ubiquitous, Pervasive and Internet Mapping, Tokyo Japan, September 7-9, pp. 9-13

[4] Morris, C., 1960, "Foundations of the theory of signs", International encyclopedia of unified science; Vol. 1: Foundation of the unity of science, No. 2, Univ. of Chicago Press, 59 pages

[5] Proceedings of the First International Joint Workshop on Ubiquitous, Pervasive and Internet Mapping, 2004, Tokyo Japan, September 7-9, 194 pages

Section II Positioning

Chapter 8	Edward Verbree, Sisi Zlatanova Positioning LBS to the third dimension
Chapter 9	Günther Retscher Altitude Determination of a Pedestrian in a Multistorey Building
Chapter 10	Rudolf Pailer, Florian Wegscheider, Joachim Fabini Terminal-Centric Location Services in the IP Multimedia Subsystem
Chapter 11	Michael Thienelt, Andreas Eichhorn, Alexander Reiterer WiKaF - A Knowledge-based Kalman-Filter for Pedestrian Positioning
Chapter 12	Andreas Eichhorn Map-independent positioning of land vehicles with causative modified motion equations

Positioning LBS to the third dimension

Edward Verbree, Sisi Zlatanova

Section GIS Technology, OTB Research Institute for Housing, Urban and Mobility Studies, Delft University of Technology
e.verbree@tudelft.nl, s.zlatanova@tudelft.nl

Abstract

Computer technology has evolved to a position of being able to handle large three-dimensional (3D) data sets. The third dimension is already taken for granted for visualisation on desktop machines. All GIS/CAD vendors are offering extended 3D tools for navigation and exploration of data. NASA World Wind and Google Earth demonstrate the possibilities of 3D to all users of the Web. The advances in geoDBMS are also striking; main stream DBMS support spatial data types, which can be adapted for handling of 3D data.

In the same way, mobile computing is experiencing a similar evolution. Large numbers of 3D computer games are already available for hand-helds, which give new opportunities to geo-specialists of having 3D geo-information on mobile devices. Despite the fact that 3D mobile hardware and software technologies are currently still behind desktop 3D in terms of capabilities, the expectations are for two or three times faster development of the mobile 3D market compared to its desktop counterpart.

Since location-based services (LBS) are among the first applications that naturally should consider the third dimension, we are going to investigate and evaluate the possible options for evolving to 3D LBS. The paper is going to concentrate and analyse on all the aspects of LBS: positioning, protocols, data retrieval and visualisation. Aspects typical for 3D data sets (large amounts of data, texturing, representations, data

models) will be discussed in details. The paper will conclude with recommended topics for research and developments.

1 Introduction

Location-based Services are often referred to as utterly location-responsive Geographical Information Systems, with the aid of location-sensitive devices and location-aware services. Compared to mobile GIS, LBS is distinguished by considering the location of the mobile user in order to serve him or her with some added value information adaptive to that location. In this sense, these services are triggered to respond at certain locations according a specific user profile. As the user can be anywhere, the position fix should not be limited; here we will discuss LBS anytime, everywhere. The position determination should then be performed in full 3D space, thus not restricted only to outdoors and ground level applications.

In this chapter, we focus on the possibilities and limitations in 3D positioning with respect to LBS. After a short introduction to general available 3D positioning systems, the possibilities of combined GPS-Galileo positioning in urban environments will be addressed by a simulation study on the availability of sufficient signal reception within city areas. This chapter concludes with a discussion on required research and developments.

2 The need for 3D LBS anytime, everywhere

Location-based Services are currently in use in many different areas and applications. The Location Services (OpenLS) of the Open Geospatial Consortium (OGC) comprise therefore an open platform for position access and location-based applications targeting mobile terminals. The OpenLS feature set is defined by the "Core Services and Abstract Data Types (ADT)". The most important aspect in these specifications is the ‘GeoMobility Server’, that provides requested information considering the position of the user. A Location Service Client sends a request for a position determination to the Gateway. The Gateway determines the position of the subscriber’s mobile terminal and forwards this to the Location Service Client.

The minimum number of OpenLS services is currently defined as directory, route, location utility (geocoding and reverse geocoding), gateway and presentation (see Figure 1). These five core services are considered sufficient (Togt et.al., 2005) for a variety of use cases such as:

- Proximity: find something in a given area;
- Fencing: restrict the position of a user to within (our just outside) a given area;
- Navigation (compute route);
- Tracking ('record' the way of a user).

These services are, however, limited to 2D, as the position ADT is restricted to latitude and longitude only.

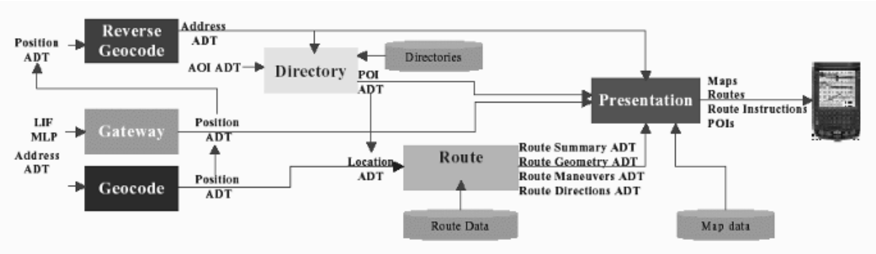


Fig. 1. GeoMobility Server (according to OpenLS)

3D LBS have to be able to ensure the same set of services, i.e., proximity, fencing, routing and tracking but in 3D (Zlatanova, et.al., 2005). Example of 3D requests would be:

- Proximity: 'show to me all the electrical switches for all floors in a building';
- Fencing: 'tell me when I am not at a dangerous floor of a building '
- Routing: 'compute a route from this floor to the ground level';
- Tracking: 'track this visitor all the way on his route from this floor to the ground level'.

In terms of core services as specified in OpenLS, 3D LBS needs to provide:

- 3D location utility service, i.e., 3D geo-coding and 3D reverse geo-coding to transform a 3D position to a descriptive location and visa-versa;
- 3D gateway service: to fetch a 3D position;
- 3D route service, i.e., give the route in multilevel constructions (buildings, viaducts, bridges, etc.);

- 3D directory, access to an online directory to find the nearest or a specific 3D place, product or service;
- 3D presentation, i.e., 3D visualisation on mobile, hand-held devices and the appropriate interface for this

To be operational, these core services should have to work with a 3D position ADT, thus latitude, longitude, and height (related to some reference system). However, with respect to Location-based Services within the built environment, 3D positioning of location-sensitive devices anytime, everywhere, is still a challenge (Zlatanova et.al., 2003).

This issue is part of the developments of the OGC Web Services, phase 3 (OGC, 2006). Herein it is concluded that: “The present OpenLS services and information model are limited to outdoor navigation (i.e., the concepts of ‘location’ and ‘navigation’ are confined to outdoor activities.) An enhancement to the OpenLS services and information model is to support seamless indoor-outdoor navigation. The OpenLS services and information model will have to be modified to accommodate indoor location and navigation constructs”. Therefore it is suggested to: “add a tracking service that supplies a position management and access capability and make first steps toward path-planning and navigation in buildings and other environments beyond the limits of road networks”. By this kind of indoor-outdoor navigation is should be made possible for “clients with mobile terminals to receive location and navigation guidance indoors or outdoors, as well as receive navigation guidance across indoor-outdoor and indoor-indoor transition points (e.g., doorways)”. Indoor location concepts must be supported for how people identify location for indoor environments, e.g., building, floor, room, etc. Indoor navigation concepts must also be supported for how people negotiate their way around indoor environments, e.g., ‘park on level P1-P4’, ‘elevator to 3rd floor’, ‘right hall to room 310’.

It should be noticed, however, that the OpenLS specifications suggest a Gateway service that uses telecommunication networks for localization of the mobile user. Positioning with mobile communication networks is widely used for commercial, push LBS applications as it is quite easy to reach a group of cellular phone users within the area of a certain base station (Cell of Origin) and send them an advertisement SMS. It is also possible to position the users within a certain sector and range of the base station by Uplink Time Difference of Arrival. If that information is monitored over time and combined with a road network, the position of the cellular phone user can be detected in a more precise manner. For example, LogicaCMG has introduced the so-called Mobile Traffic Service where these locations are aggregated to real-time traffic information for the Dutch Province of Brabant (LogicaCMG, 2005).

However, positioning with telecommunication networks based on just cell-id identification is inaccurate. Precise Mobile Network positioning requires considerable modifications to the current GSM network setup, or the use of next generation networks like UMTS. Due to the more or less planar arrangement of the GSM/UMTS beacons, accurate and reliable 3D positioning by mobile networks is not possible. Therefore, in the next sections we concentrate on other technologies for 3D positioning, focusing on their limitations.

3 Limitations in 3D positioning technologies

People can be positioned by a variety of other means such as Global Navigation Satellite Systems (GNSS) as GPS, location fingerprinting based on Wireless Local Area Network (WLAN), tags based on RFID, and other tracking approaches (Zlatanova et al, 2004). The location can also be given by using other non-coordinate related approaches such as address, floor, room, or a description of the environment. The location can be given (supposed the end-user posses the device/tool that would allow him to position himself) or requested (obtained by the system) by a user or by a control facility, for instance to locate a person in a building. Clearly, the localisation is one of the most important steps in providing any services. In this section, we will stress on the limitations of several 3D positioning approaches.

Theoretically, obtaining 3D coordinates at global scale is available. GPS devices, and other receivers to Global Navigation Satellite Systems (GNSS) like the in near future available Galileo can compute either Cartesian (X,Y,Z) or ellipsoidal (latitude, longitude, height) WGS84 coordinates. In multi-level 3D structures, the problem may come from two sides: geo-coding of the height and availability of sufficient satellites. The altitude is given as the distance to the WGS84 rotational ellipsoid. It is difficult to be linked to expressions used in daily life by references like ‘on/under the bridge’, ‘floor’, ‘base’, ‘ceiling’, ‘top’, etc. It is well known that a GNSS receiver cannot work inside or at other places with poor satellite coverage, i.e. less than four satellites in line-of-sight. Many systems exist that claim to solve that problem by applying another type of measurement technique. If it is not possible to detect enough GPS satellites in line-of-sight, some close-range pseudolites transmitters could be used also. For example, the company Noviriant offers the so-called Teralite XPS systems, a single frequency ground-based signal generator broadcasting XPS signals to mobile GPS+XPS receivers (Noviriant, 2006).

For indoor applications, a more dedicated pseudolite-only setup could be used (Kee, 2001). However, if the user is free to move in height, the transmitters should be arranged in a 3D surround setting to obtain a reliable 3D position fix.

Assisted GPS (AGPS) combines the better of two worlds: GPS and Mobile Networks (Goran et al, 2001). When the Geo-Mobility Server requests a location, the wireless network sends the approximate location of the handset (generally the location of the closest cell site) to the location server. The location server then tells the handset which GPS satellites should be relevant for calculating its position. The handset takes a reading of the proper GPS signals, calculates its distance from all satellites in view and sends this information back to the location server to let its position be determined. In hard conditions like inside locations, it is still difficult or even impossible to ‘see’ enough satellites and thus to obtain a position fix. Moreover, inside conditions and urban canyons are known to have multi-path problems, causing unreliable pseudo-ranges and thus fault determined positions.

In the last couple of years, location fingerprinting techniques using WLAN have been suggested for indoor areas where GPS does not work well (Keamarungsi et al, 2004, Xiang et al, 2004). Generally, the deployment of fingerprinting based positioning systems can be divided into two phases. First, in the offline phase, the location fingerprints are collected by performing a site-survey of the received signal strength (RSS) from multiple Access Points (APs). The RSS is measured with enough statistics to create a database or a table of predetermined RSS values. The vector of RSS values at a certain location is called the location fingerprint of that location. Second, in the online phase, a Mobile Station (MS) will report a sample measured vector of RSS from different APs to a central server. The server uses an algorithm, i.e. the Euclidian distance between the measured MSS vector and each fingerprint in the database, to estimate the location of the Mobile Station.

One major limitation of WLAN fingerprinting is, compared to GPS, the not controlled ‘space segment’, thus the locations and the signals strength of the WLAN Access Points. As these APs are not originated to be used as position beacons, they can be freely moved, rotated or in other ways disturbed from their original settings (i.e. switched off), causing a non-valid location fingerprint. The estimation of the location in the online phase could be wrong, without any possibility to detect that. Although WLAN location fingerprinting could result in accurate results, they are not reliable. A second limitation has to be found in the mapping of the RSS values. For 3D location fingerprinting this mapping has to be performed in 3D space, thus sampling of fingerprints covering the 3D space.

4 Opportunities of combined GPS-Galileo positioning for urban environments

Compared to other positioning sensors for LBS, for instance using radio-signals for mobile communication, a GNSS has the advantage of offering worldwide coverage. On the other hand, its weakest property is the requirement of, in principle, direct lines of sight to the transmitting satellites, which can be hard to realize particularly in urban environment, where typically most of the LBS applications will be used (Verbree et al, 2004).

One cannot always take measurements to determine whether a GNSS is available within urban areas. One has to realize that visibility of the satellites is not only determined by the location of the observer and obstructions around him, but also by the moment of observation as the satellites are in orbit. Besides, the actual observation of the availability of GPS during a day at or nearby a busy road crossing is impossible at all because of the traffic.

Simulation is the answer to these limitations. However, simulation requests a proper representation of the reality, both of the space segment as for the earth’s surface. The actual orbits of the GPS are known by the almanac, but in comparing Galileo and GPS the nominal constellation of GPS is put side by side with the (proposed) orbits of Galileo. We have calculated the elevation and azimuth angles for the 24 GPS and the 27 Galileo satellites for each minute during a full daytime for the test area in Delft, at 52 degrees northern latitude in the Netherlands. At a fixed location on earth, the geometry of both GPS and Galileo repeat after approximate 24 hours.

The old city of Delft has very narrow streets with built-up areas of around 8-10 meters, with famous Dutch roof shapes. A partial area of the city is modeled in three dimensions by airborne laser-altimetry in combination with the cadastral data of the parcel boundaries. The 3D city model is built up by a polygonal representation of the canals, the streets and the roofs. The quaysides and the walls - the connections between the streets and the canals at one hand and the connection between streets and the roof tops at the other - are thought to be vertical and modeled as vertical polygons. The visibility calculation however is based on a Triangulated Irregular Network (TIN) that does not allow vertical polygon constrains. The solution to that problem is found in a buffering of both roofs and canals polygons by minus 10 centimeters. These datasets are the input for the surface model of this part of Delft represented by a TIN. A height-rendered image of this TIN is shown in figure 2.

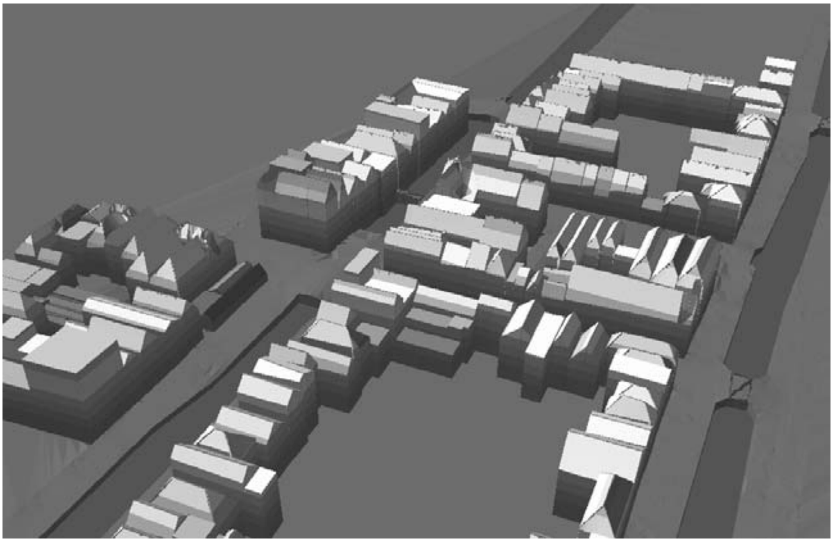


Fig. 2. 3D city model of Delft (TIN, height-rendered)

The actual visibility calculation is performed within the GIS package ArcView 3.2a with the extension 3D Analyst. The high-level scripting language Avenue allows fast prototyping of the algorithm with a proper visual feedback of the results within both 2D and 3D scenes. The simulation of the availability of GNSS consists of two algorithms. The first one calculates the total of targets (satellite positions at a certain time) seen from the observation point by:

```
Count = 0
for each aTarget 'possible observable satellite
  if (aTIN.LineOfSightsAsShapes (anObserver, aTarget,
      ListOfShapes) = True) then
    Count = Count + 1
  end
end
```

Satellite signal propagation is assumed to take place along geometric straight lines. The request `aTIN.LineOfSightsAsShapes` returns not only whether or not `aTarget` is visible from `anObserver` across `aTIN Surface`, but also returns `aListOfShapes` containing `theObstaclePointZ` and the visible or invisible parts of the profile line. See figure 3 for the visual feedback of this calculation, with red the GPS visibility and within purple the Galileo visibility.

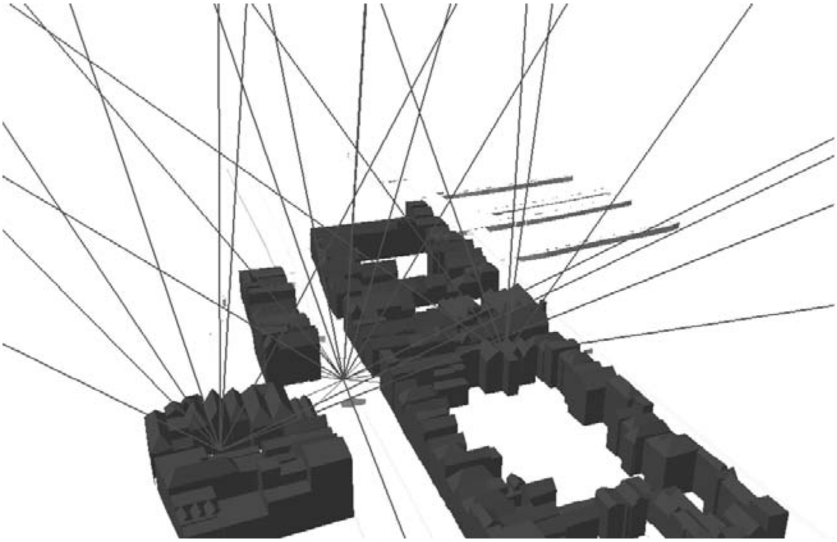


Fig. 3. Some examples of combined visibility of GPS and Galileo

The second algorithm calculates the availability of 'enough' satellites during a day time. We have chosen about 50 test observer points, with a height of 1.80 m above street level. For each of these observer points the total number of visible satellites during a day time is calculated. During a day time means 60*24 = 1440 different constellations for both GPS and Galileo. Each target (24 GPS satellites and 27 Galileo satellites) within these constellations is checked by the request: `aTIN.ObscuresTarget (anObserver, aTarget).`

For GPS and Galileo alone observing four satellites simultaneously is the minimum for a position fix, without any preliminary knowledge as a known height. In combination of GPS and Galileo this requirement for sufficient availability holds true, but this demand can be extended with at least three GPS and two Galileo or two GPS and three Galileo satellites. Availability is analyzed here regardless the actual geometry of the visible satellites, which can have a large impact on the eventual position accuracy. The percentage of 'valid' cases gives an indication of the availability of GPS, Galileo and the combination of both within urban areas.

From this analysis, it can be concluded that the coverage of GPS alone is not sufficient within urban areas (i.e. less than 50% availability in narrow streets). However, for navigation purposes it has to be stated that the visibility on street crossings is far better than within street lanes. However, decisions where to go are made at crossings and car navigations systems use map-matching and auxiliary sensors to keep the car on track. A second consideration should be made upon the required visibility of four satellites. If the height is known and steady, (as in the streets) a position fix can be calculated out of the measurements to three visible satellites. This will improve the availability map considerably.

The calculated availability for Galileo alone is better than for GPS. The amount of proposed satellites (27 for Galileo compared to 24 for GPS) is due to this result. Again, the results will improve when relaxing the demand of four visible satellites to three.

The combination of GPS and Galileo is very promising, which is not surprisingly with 51 satellites to choose. Only half of them are above the horizon, but it is clearly shown that - besides very narrow streets - the availability of the combination is nearly 100%. This result is however a little optimistic, because we have not taken into account the obstruction by trees and obstructions other than buildings. As a compromise, it is to be noted that for both Galileo and GPS the plain nominal satellite constellation was used.

As discussed in section 3, WLAN positioning is based on a kind of fingerprinting where providers have to generate first and foremost a database of signal characteristics of received radio waves at various, but known, locations. Once known, the position of the devices is determined by matching this database of digital fingerprints with the incoming signal. A basic form of fingerprinting could be applied to GNSS, as one of the main features of GNSS is the necessity of sufficient free line-of-sights to the satellites. As this characteristic can be determined within built environments in forehand, by considering a 3D city model, a GNSS fingerprint could be calculated. This GNSS fingerprint could increase the usability of GNSS within the built environment considerable.

5 Discussion

The general conclusion is that sufficient technological possibilities for 3D positioning exist but all of them do have their drawbacks with respect to Location-based Services, i.e., coverage (indoor/outdoor), availability (not anytime up and running), precision, reliability and integrity.

Most position systems are presented as ‘stand-alone’ solutions, due to commercial interests. As no system operates best under all circumstances, the reliability will be improved and ensured when the systems become more integrated.

The OpenLS specification has to be further extended to provide 3D core services:

- 3D location utility service, i.e. 3D geo-coding and 3D reverse geo-coding to transform a 3D position to a descriptive location and visa-versa;
- 3D gateway service: to fetch a 3D position;
- 3D route service, i.e. give the route in multilevel constructions (buildings, viaducts, bridges, etc.);
- 3D directory, access to an online directory to find the nearest or a specific 3D place, product or service;
- 3D presentation, i.e. 3D visualisation on mobile, hand-held devices and the appropriate interface for this.

In this respect, these core services should have to work with a 3D position ADT, thus latitude, longitude, and height (related to some reference system).

Presently, all 3D positioning techniques have particular limitations, but many developments are in progress and hopefully will take place in the coming few years. For example, 3D localisation based on a combined GPS and Galileo constellation – in conjunction with GNSS-fingerprinting - is expected to resolve some of the typical urban canyon problems, despite the performed calculation are a bit idealistic. When Galileo becomes operational, real tests should conclude on the effect of obstructing trees and other objects.

References

Goran M. Djuknic and Robert E. Richton, 2001, *Geolocation and Assisted GPS*, Computer, vol. 34, no. 2, pp. 123-125, February, 2001.

Kaemarungsi, Kamol and Prashant Krishnamurthy, 2004, *Modeling of Indoor Positioning Systems Based on Location Fingerprinting*, IEEE INFOCOM, 2004, 11 p.

Kee, Changdon, 2001, *Centimeter-Accuracy Indoor Navigation Using GPS-Like Pseudolites*, GPS World, November 2001.

LogicaCMG, 2006. www.logica-cmg.com; www.mts-live.com.

Noviriant, 2006. www.novariant.com/mining.

OGC, 2006. Open Geospatial Consortium: www.opengeospatial.org

Togt, R., E. Beinat, S. Zlatanova and H.J. Scholten, 2005, *Location interoperability services for medical emergency operations during disasters,* in: P.J.M. van Oosterom, S. Zlatanova & E.M. Fendel (Eds.), Geo-information for disaster management, Springer Verlag, Heidelberg, pp. 1127-1141.

Verbree, Edward, Christiaan Tiberius and George Vosselman, 2004, *Combined GPS-Galileo positioning for Location Based Services in urban environment*, in: Proceedings of the Symposium on Location Based Services & Telecartography, 28-29 January, 2004, Vienna, Austria, pp 99-107.

Xiang, Z. et.al., 2004, *A wireless LAN based positioning system*, IBM Journal Res&Dev, Vol 48, Sep./Nov., 2004, pp. 617-626.

Zlatanova, S. and E. Verbree, 2003, *Technological Developments within 3D Location-based Services*, International Symposium and Exhibition on Geoinformation 2003 (invited paper), 13-14 October 2003, Shah Alam, Malaysia, pp. 153-160.

Zlatanova, S. and E. Verbree, 2004, *User tracking as an alternative positioning technique for LBS*, in: Proceedings of the Symposium on Location Based Services & Telecartography, 28-29 January, 2004, Vienna, Austria, pp. 109-115.

Zlatanova, S. and E. Verbree, 2005, *The third dimension in LBS: the steps to go*, in: Proceedings of the 3rd Symposium on Location Based Services & TeleCartography, 28-30 November 2005, Vienna, Austria, pp. 185-190.

Altitude Determination of a Pedestrian in a Multi-storey Building

Günther Retscher

Institute of Geodesy and Geophysics, Vienna University of Technology, Gusshausstrasse 27-29, A – 1040 Wien, Austria, gretsch@pop.tuwien.ac.at

Abstract

A challenging task in using pedestrian navigation and guidance services indoors is to determine the correct floor of a user in a multi-storey building because most indoor location techniques only provide 2-D information. In this case it can be recommended to augment the position determination system with a barometric pressure sensor for direct observation of height differences. In the research project NAVIO (Pedestrian **Nav**igation Systems in Combined **I**ndoor/**O**utdoor Environments) tests with different sensors have been performed and their results are presented. The tests show that it is possible to determine the correct floor of a user using a barometric pressure sensor as the standard deviation of the estimation of the height differences is better than ± 1 m.

1 Introduction

In recent years new technologies and methods for positioning in indoor environments have been developed. Useable geolocation techniques include cellular phone positioning, the use of WiFi or WLAN (wireless local area networks), UWB (ultra-wide band), RFID (radio frequency identification), Bluetooth and other systems using infrared, ultrasonic and radio signals (see e.g., Retscher, 2005b). In this chapter a brief overview of some of these technologies that can be employed for personal navigation systems is given. Most of the systems, however, are able to locate the user only in two dimensions and the altitude (or height) of the user is not determined. Then the height has to be observed using an additional sensor to be able to locate a user on the correct floor of a multi-storey building. In the research project NAVIO (Gartner, et.al., 2004) at our University the use of a barometric pressure sensor for the direct observation of height differences is suggested. Test measurements performed in our 5-storey office building are presented.

2 Overview of indoor location systems

For indoor positioning different location techniques have been developed which use signals such as infra red, ultra sonic, radio signals or visible light (Retscher and Kistenich, 2006). Methods for position determination include Cell of Origin (CoO) where the location of the user is described in a certain cell area around the transmitter, Time of Arrival (ToA) where the travel time of a signal between a transmitter and receiver is obtained, Time Difference of Arrival (TDoA) where the time difference of signals sent from a transmitter is determined at two receiving stations, signal strength measurement for location determination using fingerprinting (e.g. WiFi or WLAN fingerprint, see Retscher, 2004) where the signal strength values are compared with previous stored values in a database and the location of the user is obtained using a matching approach, and location determination using digital images (Retscher and Kistenich, 2006). Table 1 gives an overview about different indoor location techniques.

The systems Active Badge (Want et al., 1992) and WIPS (Roth, 2004) employ infra red signals for location determination, Active Bat (Hightower und Boriello, 2001) and Cricket (Roth, 2004) use ultra sonic signals. For the location of cellular phones ToA or TDoA measurements can be performed (Retscher, 2002). Satellite or similar signals are also employed for the location of cellular phones using Assisted GPS (A-GPS) or for the Australian system Locata (Barnes et al., 2003) which makes use of standard RTK positioning with GPS similar signals. For indoor positioning the use of WiFi or WLAN (Wireless Local Area Networks) has become popular and the systems Radar (Bahl and Padmanabhan, 2000), IMST ipos (Imst, 2004), Ekahau (Ekahau, 2005) and WhereNet (WhereNet, 2005) are examples. Apart from WiFi also Ultra Wide Band (UWB) signals and Bluetooth (Hallberg et al., 2003) can be employed. SpotON employs also

Table 1. Comparison of indoor location techniques

System name	Signal	Method	Absolute Positioning	Relative Positioning	Positioning	Tracking	Geometrical	Symbolic	Costs	Positioning Accuracy [m]
Active Badge	IR	CoO	✓			✓		✓	low	room
WIPS	IR	CoO	✓		✓			✓	low	room
Active Bat	US	ToA	✓			✓	✓		low	0,1
Cricket	US	ToA	✓	✓	✓			✓	low	1,2
GSM	RS	TDoA AoA	✓			✓	✓		low	50-100
A-GPS	RS	ToA	✓		✓		✓		high	20-25
Locata	RS	ToA	✓		✓		✓		high	0,1-1
Radar	RS	SS	✓		✓	✓	✓		high	3-4
IMST ipos	RS	SS	✓			✓	✓		high	1-3
Ekahau	RS	SS	✓		N/A	N/A	✓		high	1-3
WhereNet	RS	SS	✓		N/A	N/A	✓		N/A	N/A
UWB	RS	ToA TDoA	✓		✓		✓		high	0,2
Bluetooth	RS	CoO	✓		✓	✓	✓		average	10
SpotON	RS	SS	✓	✓	✓	✓	✓		average	1 m^3
RFID	RS	CoO		✓		✓		✓	low	1-20
CyberCode	VL	DI		✓	✓			✓	average	variable
Ubitrack	VL	DI		✓		✓	✓		N/A	N/A
EasyLiving	VL	DI	✓			✓	✓		high	variable

The following abbreviations are used in Tabel 1:

Signals:
IR.......Infra red
US......Ultra sonic
RS......Radio signals

Postioning Methods:
CoO.... Cell of Origin
ToA.... Time of Arrival
TDoA.. Time Difference of Arrival

VL......Visible light	AoA.... Angle of Arrival
	SS….. Signal strenght measurement
N/A.....not available	DI…… Digital images

radio signals and performs signal strength measurements (Hightower et al., 2000). Table 1 also contains three systems using digital images for location determination, i.e., CyberCode (Rekimoto et al., 2000), Ubitrack (Newman et al., 2004) and EasyLiving (Brumitt et al., 2000).

For navigation and way finding in smart environments the use of RFID (Radio Frequency Identification) for ubiquitous positioning is also a promising solution. RFID is a method of remotely storing and retrieving data using devices called RFID tags. An RFID tag is a small object, such as an adhesive sticker, that can be attached to or incorporated into a product. RFID tags contain antennas to enable them to receive and respond to radio-frequency queries from a RFID transceiver. For location determination RFID tags can be placed on active landmarks or on known locations in the surrounding environment. If the user passes by with a RFID reader, the tag ID and additional information (e.g. the 3-D coordinates of the tag) are retrieved. Thereby the range between the tag and reader in which a connection between the two devices can be established depends on the type of tag. RFID tags can be either active or passive. Passive RFID tags do not have their own power supply and the read range is less than for active tags. They have practical read ranges that vary from about 10 mm up to about 5 m. Active RFID tags, on the other hand, must have a power source, and may have longer ranges and larger memories than passive tags, as well as the ability to store additional information sent by the transceiver. At present, the smallest active tags are about the size of a coin. Many active tags have practical ranges of tens of metres, and a battery life of up to several years. The location method is Cell of Origin (CoO) and the size of the cell is defined by the range of the tags. Therefore using active RFID tags the positioning accuracy ranges between a few metres up to tens of metres and with passive tags up to about 5 m. Although this positioning accuracy can be low for some applications, RFID positioning can be very useful in combination with other sensors.

3 Altitude determination of a pedestrian using a barometric pressure sensor

In the research project NAVIO the Vaisala pressure sensor PTB220A (Vaisala, 2005) is employed for the determination of height differences from changes of the air pressure. The PTB220A is designed for measurements in a wide environmental pressure and temperature range with an extremely high accuracy. Starting from a given height the pressure changes can be converted in changes in height using the following equation:

$$\Delta H = H_2 - H_1 = 18464 \cdot (1 + 0{,}0037 \cdot t_m) \cdot (\lg B_1 - \lg B_2)$$

where ΔH is the height difference between two stations 1 and 2, B_1 and B_2 are the pressure observations at station 1 and 2 and t_m is the mean value of the temperature of both stations. It must be noted that this equation is an approximation formula that is valid for central Europe only (Kahmen, 1997). Tests showed that there is no significant difference between the results using the approximation formula and an equation derived from Jordan which is also valid for other parts in the world and takes into account the geographic location of the two stations.

4 Sensor tests: locating the user on the correct floor in a multi-storey building

Substantial sensor testing has shown that we are able to determine the correct floor of a user in a multi-storey building using this sensor. First of all the drift of the sensor was analyzed in several long term tests. Then it was investigated, if a functional connection between the observed pressure differences and the height differences can be derived. This connection can be described using characteristic curves. Finally the accuracy of the height determination was analyzed in detail.

Figure 1 shows the test area which is in our office building at the Vienna University of Technology. The trajectory leads from the main entrance of the building via two staircases to our institute which is located on the third floor of the 5-storey building.

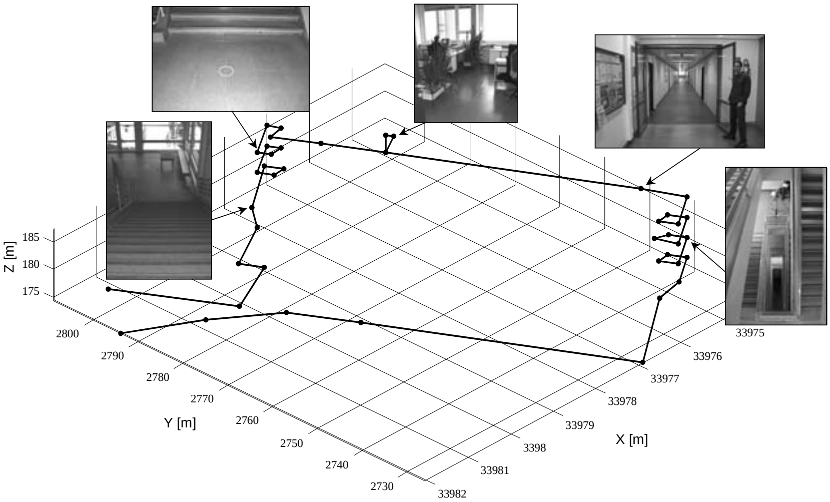


Fig. 1. Indoor trajectory in our office building at the Vienna University of Technology from the main entrance onwards to our institute located on the third floor of the building

5 Determinations of the sensor drift

Figure 2 shows an observation of the sensor over two hours performed on a benchmark located on the roof of our building. The height of the benchmark is 200.05 m. The sampling rate of the observations was one second. During the two hour observation period the temperature dropped from 16.5° C to 15.8° C. The maximum deviations from the given height reach values of + 0.42 m and – 0.64 m in the first half hour. Thereby the observations can vary randomly in this range of about 1.06 m. Some discontinuous variations are also caused by the wind during the observation period. In summary, it can be concluded that no significant drift rates could be seen during several long term tests. The influence of wind, temperature changes and the air conditioning system inside the building, however, affects the results and can be clearly seen in the observations. For the first half hour of operation variations of the air pressure in the range of ± 0.15 hPa could be seen. Considering the resolution of the sensor which is 0.01 hPa, the sensor can be regarded as stable and no drift rate will be considered in the following.

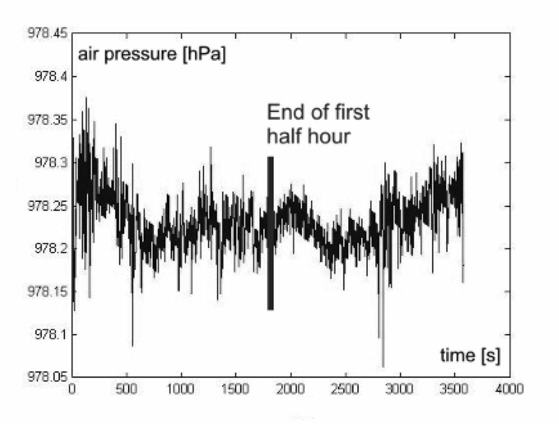


Fig. 2. Long term sensor observations with the Vaisala pressure sensor PTB220A on benchmark No. 11 on the roof of our building

6 Determination of a characteristic curve for the barometric pressure sensor

A characteristic curve shows a functional connection between the observed pressure differences and the height differences. If such a curve exists and the functional connection is linear, then the pressure differences can be converted into height differences. For this purpose observations in the building have been carried out during different times of the day. Figure 3 shows six observations in the morning of one day and the resulting linear characteristic curve. The start point of three of these observations was on the ground floor and for the other three on the roof of the building. The characteristic curve is given by the following equation:

$$\Delta h = 8.769 \cdot \Delta p$$

where Δh is the height difference and where Δp is the difference in air pressure.

As can be seen in Figure 3, the measurement series show a good agreement with the resulting characteristic curve. There is also no difference, if the observations begin either on the ground floor or on the roof of the building. Further characteristic curves have been obtained from different measurement series and can be found in Retscher & Kistenich (2005). As a result it can be seen that we are able to calculate a linear characteristic curve which describes a linear functional connection between the air pressure observations and the changes in height.

7 Determination of the height in the building

Figure 4 shows observations with the Vaisala pressure sensor PTB220A in our office building starting from the main entrance up to the third floor of the building where our institute is located. It can be clearly seen that the sensor is able to determine the correct floor of the user with a high precision. The standard deviation of the pressure observation is in the range of ± 0.2 hPa and the maximum deviation of the determined height is less than ± 1 m for 91 % of the observations. Thereby the deviations depend also on the time of day; higher deviations are obtained during noon where usually more people are inside the building and higher variations of the air pressure occur caused by higher air circulation due to frequent opening of doors and windows. The maximum outlier during noon reaches values of about 1.4 m. In summary, it can be concluded that the sensor is able to locate the user on the correct floor.

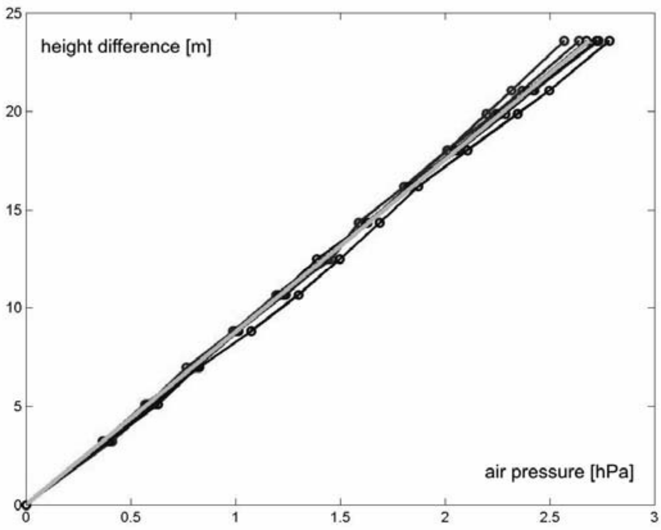


Fig. 3. Linear characteristic curve of six observations in our office building between the ground floor and the roof (6th floor)

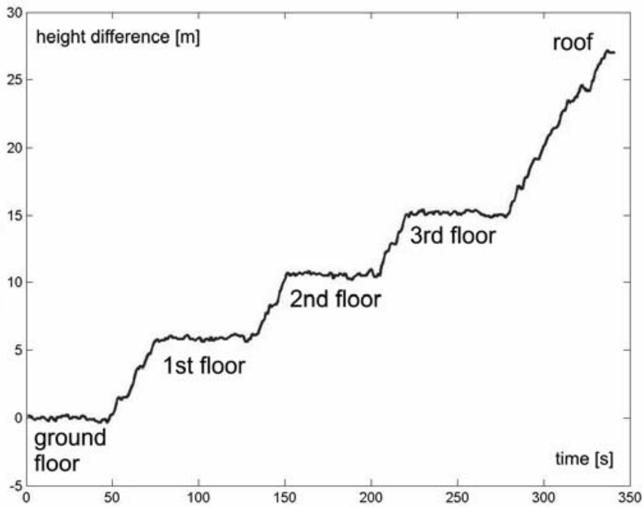


Fig. 4. Test measurements with the Vaisala pressure sensor PTB220A in our office building at the Vienna University of Technology

8 Concluding remarks and outlook

A variety of systems have been developed in recent years to determine the location of people and objects within building. This chapter has provided an overview about their operation and performance. They provide mostly only 2-D location determination of the user and the determination of the correct floor is a very challenging task. Therefore, in the research project NAVIO, the use of a barometric pressure sensor for direct observations of the altitude is investigated. It be can seen from the presented sensor tests that the elevation of a pedestrian within a building can be determined with high precision and reliability if a barometric pressure sensor is employed. Then, in combination with other indoor location techniques and dead reckoning sensors, a continuous 3-D position determination in indoor environments is possible. Recently an indoor location system based on WiFi fingerprinting has been installed in our office building and we are working on its integration in our multi-sensor system. Using the WiFi system it is possible to locate the user in two dimensions with a positioning accuracy in the range of 1 to 3 m. The system has been tested in a study and the interested reader is referred to (Retscher et al., 2006).

In our multi-sensor approach the observations of all sensors have to be combined and an integrated position solution has to be obtained. For the integration of all observations a multi-sensor fusion model based on an extended Kalman filter which makes use of a knowledge-based pre-processing of the sensor observations shall be applied. The principle of this approach is presented in Retscher (2005a). The fusion model consists of two steps, i.e., a knowledge-based pre-processing filter followed by a central Kalman filter for optimal estimation of the current user's 3-D position. The knowledge-based pre-processing filter represents an extension of common multi-sensor fusion models in a way that the data based system analysis and modelling is supplemented by a knowledge-based component and therefore not directly quantifiable information is implemented through formulation and application of rules. These rules are tested in the pre-processing step and if they are fulfilled certain actions are executed. Due to the knowledge-based analysis of the sensor observations gross errors and outliers can be detected and eliminated in this processing step. In addition, the pre-processing filter supplies input values for the stochastic model of the central Kalman filter. Therefore, the weightings of the sensor observations can be adjusted in the Kalman filter depending on the availability and quality of the current observations. This integration approach will be implemented and further sensor tests will be carried out to test and analyze this approach. Due to the development of new advanced sensors it can be expected that multi-sensor solutions which provide location capabilities in outdoor and indoor environments will be deployed in pedestrian's navigation services in the near future. We believe that these services will play an important role in the field of location-based services.

Acknowledgements

The research work presented in this paper is supported by the FWF Project NAVIO of the Austrian Science Fund (Fonds zur Förderung wissenschaflicher Forschung) (Project No. P16277-N04). The author would like to thank Mr. Michael Thienelt and Mr. Michael Kistenich for the performance of test measurements with the barometric pressure sensors and the preparation of some of the figures in this paper.

References

Bahl, P. and Padmanabhan, V. (2000) Radar: An In-Building RF-based User Location and Tracking System, Proceedings of IEEE Infocom, Tel-Aviv, pp. 775-784.

Barnes, J., Rizos, C. and Wang, J. (2003) Locata: A New Positioning Technology for High Precision Indoor and Outdoor Positioning, in: Papers presented at the ION GPS/GNSS 2003 Conference, September 9-12, 2004, Portland, Oregon, USA, unpaginated CD-Rom Proceedings.

Brumitt, B., Meyers, B., Krumm, J., Kern, A., and Shafer, S. (2000) EasyLiving: Technologies for Intelligent Environments, Microsoft Corporation.

Ekahau (2005) http://www.ekahau.com, Last date accessed 07.2005.

Gartner, G., Frank, A. and Retscher, G. (2004) Pedestrian Navigation System for Mixed Indoor/Outdoor Environments. in: Gartner G. (Ed.): Geowissenschaftliche Mitteilungen, TU Wien, Heft 66, 2004, pp. 161-167.

Hallberg, J., Nilsson, M. and Synnes, K. (2003) Bluetooth Positioning, Proceddings of the 10th International Conference on Telecommunications (ICT), Tahiti, Feb. 23 – March 1.

Hightower, J., Want, R. and Boriello, G. (2000) SpotON: An Indoor 3D Location Sensing Technology Based on RF Signal Strength. University of Washington, Department of Computer Science and Engineering.

Hightower, J. and Boriello, G. (2001) A Survey and Taxonomy of Location Systems for Ubiquitous Computing, Technical Report, University of Washington, Department of Computer Science and Engineering.

Imst (2004) Indoor Locating – Imst ipos, Project c21, Presentation, IMST GmbH, Carl-Friedrich-Gauß-Str. 2, D-47475 Kamp-Lintfort, Germany, see also http://www.centrum21.de/, Last date accessed 07.2005.

Kahmen, H. (1997) Vermessungskunde. 19. Auflage, Walter de Gruyter, Belin, New York, pp. 475-482 (German).

Kistenich, M. (2005) Indoor Positioning: Vergleich von Systemen zur Positionsbestimmung und Navigation in Gebäuden, Diploma thesis, Institute of Geodesy and Geophysics, Vienna University of Technology.

Newman, J., Wagner, M., Bauer, M., Mac Williams, A., Pintaric, T., Beyer, D., Pustka, D., Strasser, F, Schmalstieg, D. and Klinker, G. (2004) Ubiquitous Tracking for Augmented Reality, Technical Report, Vienna University of Technology, http://www.ims.tuwien.ac.at/media/documents/publications/ ubitrackismar04.pdf, Last date accessed 07.2005.

Rekimoto, J. and Ayatsuka, Y. (2000) CyberCode: Designing AR Environments with Visual Tags, Proceedings of DARE 2000, Elsinore, pp. 1-10.

Retscher, G. (2002) Diskussion der Leistungsmerkmale von Systemen zur Positionsbestimmung mit Mobiltelefonen als Basis für LBS, in: Kelnhofer, F., Lechthaler, M., Brunner, K. (Eds.): Geowissenschaftliche Mitteilungen, TU Wien, Heft 58, pp. 41-58, 2002.

Retscher, G. (2004) Multi-sensor Systems for Pedestrian Navigation, in: Papers presented at the ION GNSS 2004 Conference, September 21-24, 2004, Long Beach, California, USA, unpaginated CD-Rom Proceedings, 12 pgs.

Retscher, G. (2005a) A Knowledge-based Klaman Filter Approach for an Intelligent Pedestrian Navigation System, in: Papers presented at the ION GNSS 2005 Conference, September 13-16, 2005, Long Beach, California, USA, unpaginated CD-Rom Proceedings, 11 pgs.

Retscher, G. (2005b) Indoor Positioning Technologies for Personal Navigation Systems, in Grün A., H. Kahmen (Eds.): Optical 3-D Measurement Techniques VII, Papers presented at the 7th Conference on Optical 3-D Measurement Techniques, October 3-5, 2005, Vienna, Austria.

Retscher, G. and Kistenich, M. (2005) Höhenbestimmung in einem Gebäude mit einem digitalen Barometer für die Navigation von Fußgängern. Allgemeine Vermessungsnachrichten (AVN), 10/2005, pp. 348-354 (German).

Retscher, G. and Kistenich, M. (2006) Vergleich von Systemen zur Positionsbestimmung und Navigation in Gebäuden, Paper accepted for publication in the Zeitschrift für Geodäsie, Geoinformation und Landmanagement (ZfV) (German).

Retscher, G., Moser, E., Vredeveld, D. and Heberling, D. (2006): Performance and Accuracy Test of the WLAN Positioning System ipos. in: Papers presented at the 3rd Workshop on Positioning, Navigation and Communication WPNC 2006, University of Hannover, Germany, March 16, 2006, Hannoversche Beiträge zur Nachrichtentechnik, Shaker Verlag.

Roth, J. (2004) Data Collection, in: Schiller, J. and Voisard A. (Eds.) (2004): Location-Based Services, Morgan Kaufmann Publishers, USA, pp. 175-205.

Vaisala (2005) PTB220 Digital Barometer, User's guide, Vaisala, Finland, http://www.vaisala.com/businessareas/instruments/servicesupport/userguides/barometricpressure/PTB220%20User%20Guide%20in%20English.pdf, Last date accessed: 07.2005.

Want, R., Hopper, A., Falcao, V., Gibbson, J. (1992) The Active Badge Location System, ACM Transactions on Information Systems, 10(1), pp. 91-102.

WhereNet (2005) WhereNet Homepage, http://www.wherenet.com, Last date accessed 07.2005.

Terminal-Centric Location Services in the IP Multimedia Subsystem

Rudolf Pailer[1], Florian Wegscheider[2], Joachim Fabini[3]

[1] mobilkom austria AG & Co KG, Vienna, Austria
[2] Telecommunications Research Center Vienna (ftw.), Vienna, Austria
[3] Vienna University of Technology, Institute of Broadband Communications, Vienna, Austria

Abstract

Location-based services in 3G and beyond networks depend on high-accuracy location data. Today's 2G and 3G access networks determine user location in a network-centric manner. The accuracy of these location mechanisms depends on the radio infrastructure near the user's geographical position and is typically limited to hundreds or thousands of meters. This precision is inadequate for typical location-based services like city maps or route planners.

In this paper we present a terminal-centric location enabler that integrates seamless with the existing IP Multimedia Subsystem presence architecture and interoperates with network-centric location mechanisms. We argue that the optimal accuracy is achieved by determining the location on the user terminal, which can be equipped with location sources like Global Positioning System (GPS) receivers. In consequence, the user terminal should be used as the primary source for location information. We extend the existing and well-known concept of presence by defining location as a type of presence information that is of interest to users. Terminal-based triggers and filters reduce the amount of traffic on the radio interface and enable scalable location architecture.

This paper describes the system architecture for a terminal-based location service enabler in the 3G IP Multimedia Subsystem (IMS).

1 Introduction

Location Based Services (LBS) are services that require information about the physical position of a user in order to provide ‘added-value’ to services in a 3rd generation (3G) network. Location data may be plain geographical coordinates, access point cell ids, civil location in form of postal addresses or more abstract definitions like ‘in the office’, ‘at home’. Example services are a map showing the user’s current location or triggering a switch of the user profile when entering a specified area.

Service enablers, defined to expose network functionality to external service providers, are becoming the cornerstones of modern service architectures defined by the Parlay Group, the Open Mobile Alliance (OMA) or in the IP Multimedia Subsystem (IMS). A Location Service Enabler is a functional entity in the network enabling value-added services to query the current position of a user or to request a trigger when a specified area is entered or left.

Many mobile and fixed network operators have already started to migrate their telecommunication networks towards an All-IP infrastructure where voice loses its dominancy and becomes just one among many services. The IP Multimedia Subsystem (IMS) [1], standardized by the 3rd Generation Partnership Project (3GPP), is the most promising candidate for replacing legacy, voice-dedicated mobile networks with an All-IP technology. As opposed to traditional IP-based networks, the IMS guarantees end-to-end Quality of Service (QoS) in the network. The IMS creates an infrastructure that enables the fast deployment of new IP-based services and flexible billing while still maintaining compatibility with existing applications. The 3GPP IMS location specifications [2] adopt a network-centric location mechanism based on the Gateway Mobile Location Centre (GMLC) as primary location source for LBS.

The accuracy of this network-centric positioning mechanism is dynamic. Depending on the radio infrastructure near the user’s geographical location the GMLC typically positions a user within an area of several thousand square meters in urban area to several square kilometers in rural area. This accuracy is not satisfactory for many LBS. High investments are required to implement network-centric location mechanism enhancements like Network Assisted GPS (A-GPS) or Idle Period Downlink- Observed Time Difference Of Arrival (IPDL-OTDOA).

We argue in this paper that the optimal source for location data is the user's terminal. If the terminal is equipped with some generic location technology (e.g. a GPS receiver), it can deliver the most accurate positioning information to the system. We thus propose a handset-based Location Service Enabler.

Many countries impose legal restrictions that regulate the processing of personal data and the protection of privacy in electronic communications. Furthermore users want to control who gets access to their location data and want to be sure that the transport in the network of such sensitive data is protected by strong security mechanisms.

Our analysis of the IMS presence system shows that the requirements regarding access authorization, encryption and privacy for presence are indeed identical to those for location. The concept of presence was introduced in instant messaging systems. Presence is information about the online status of other users. A watcher can subscribe to notifications about the state of a watched user, called 'presentity'. Classical presence information is defined as the willingness of the presentity to communicate.

We argue in this paper that location data can be regarded as a special type of presence information that we will subsequently call location presence data. The proposed architecture for an IMS Location Service Enabler is thus an extension of the existing IMS presence system. Our system design builds on the request routing, authorization, encryption and privacy mechanisms of the IMS presence enabler, but extends the existing specifications in order to support a distributed terminal-based Location Service Enabler.

A more detailed description of the proposed architecture can be found in [19] and [20].

2 The IP Multimedia Subsystem - IMS

The IP Multimedia Subsystem is an overlay network specified by the 3GPP in the last years and is currently in the roll-out phase. The IMS makes heavy use of the Session Initiation Protocol (SIP) [3] and its extensions and defines several service enablers such as voice over IP (VoIP), multimedia streaming, presence, instant messaging, push to talk, etc. Initial focus of 3GPP's IMS specification was on UMTS and GPRS networks but the concept of IMS is access-independent. I.e., IMS can use WiMax, WLAN or fixed networks in the access providing a guaranteed end-to-end QoS to customers. Whereas most UMTS applications can be realized without the IMS, the specified mechanisms promise a uniform, standardized way of handling: quality of service, charging, roaming and integration of different applications and services.

The core IMS network uses three component classes: the Call-Session Control Functions (CSCF), which route SIP messages and control basic session establishment, Application Servers (AS) which are also SIP-enabled and implement services in the IMS network and finally the Home Subscriber Server (HSS), which is the user data repository in the home network (see figure 1).

The User Equipment (UE) registers with the IMS network by sending its SIP registration request to the Proxy Call-Session Control Function (P-CSCF), which was assigned to the UE by the network (DHCP, PDP context activation). The P-CSCF forwards all messages it receives from the UE to the Interrogating CSCF (I-CSCF) which serves the user's home domain. The I-CSCF contacts the HSS, selects a Serving CSCF (S-CSCF) based on the services the user has subscribed to, and forwards the registration message to the selected S-CSCF. The S-CSCF is the service access point and service dispatcher within the IMS network. It authenticates the user and registers him with the IMS network. Depending on the user profile, the S-CSCF redirects calls from or to a specific user to AS like the Presence AS shown in the figure.

For simplicity the picture does not show interfaces towards legacy application servers like CAMEL or OSA/PARLAY, as well as gateways towards legacy networks.

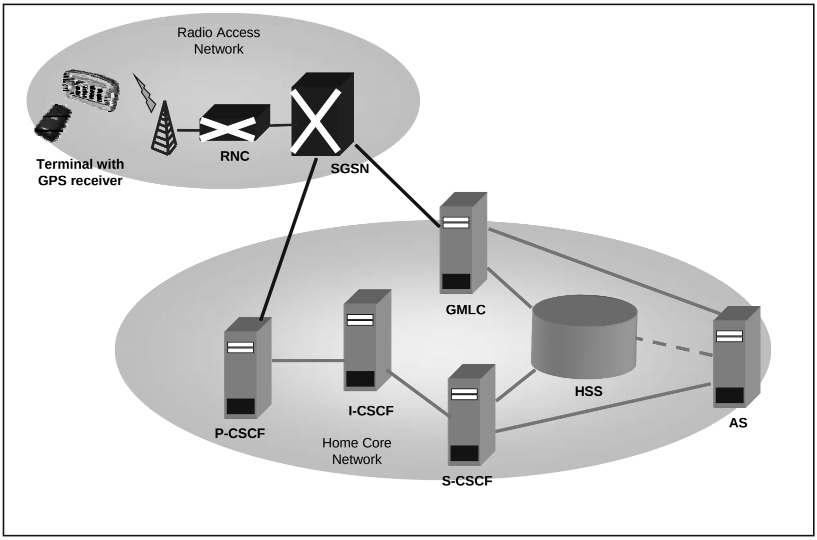


Fig. 1. IMS Components

3 Location Mechanisms and Standards

Current GSM/GPRS and UMTS networks use a network-based Location Service Enabler to provide location data to location applications. 3GPP specification TS 22.071 [4] gives a general description of location services (LCS) and service requirements for 3rd generation networks. It identifies four reasons for implementing LBS in IMS:

1. Provide value-added services to customers
2. Operators can use LBS for network optimization and OAM services.
3. Emergency services
4. Lawful intercept

3GPP TS 23.271 [5] specifies the mechanisms to support mobile location services for operators, subscribers and third party service providers. The radio signals of the wireless network are used to determine the (geographic) location of the user equipment (UE). The specification explicitly assumes that “positioning methods are Access Network specific, although commonalties should be encouraged between Access Networks”.

Interworking with the IMS has been introduced in version 6.7.0 for UMTS release 6 in March 2004. The central component in the 3GPP LCS design is the Gateway Mobile Location Center (GMLC) that offers location services.

The following network-based positioning methods are specified in the 3G documents:

- **Cell Coverage Based**. The cell ID is either known to the radio network or can be obtained by paging the terminal. The accuracy of this method depends on the cell size and is typically in the range of 300 m in urban areas.
- **Idle Period Downlink- Observed Time Difference Of Arrival (IPDL-OTDOA)**. This method measures the relative time of arrival of pilot signals from different base stations. At least three stations have to be visible to calculate a location. Network planning however optimizes a cellular network for available bandwidth which means reducing unnecessary overlapping of cells. Thus the accuracy of the IPDL-OTDOA method is not satisfying in real networks (about 50m – 150m) and the technology requires a substantial up-front investment in the radio access network.
- **Network Assisted GPS (A-GPS)**. The network sends assistance data to a GPS receiver in the terminal in order to speed up the position calculation and to provide service in areas where GPS signals are weak (e.g. inside buildings). Assistance data contains precise GPS satellite orbit and clock information, initial position and satellite selection. A-GPS offers

good position accuracy (5 m) but requires a GPS assistance service and necessitates a medium up-front investment in the 3G network. Furthermore A-GPS enabled terminals are not available on the market yet.

A terminal based location enabler on the other hand, such as a built-in GPS sensor or a Bluetooth GPS module, is a relatively inexpensive hardware feature and allows upgrading only those terminals whose users subscribe to or want to use location enabled services. A terminal based location enabler can be accessed in a peer-to-peer mode with minimal impact on network resources, thus making a highly scalable solution possible. Since a terminal centric solution is totally independent of the underlying network technology or the network provider, it enables vertical handover (e.g. UMTS - WLAN) and works in roaming scenarios.

4 IMS Presence Location Client Architecture

The concept of presence was introduced in instant messaging systems. Presence is information about the online status of other users. A watcher can subscribe to notifications about this state. The watched user is called ‘presentity’. Classical presence information is defined as the willingness of the presentity to communicate.

The presence enabler in the IMS is based on the SIP Instant Messaging and Presence Leveraging Extensions (SIMPLE [6]). The SIMPLE specifications in turn make use of the SIP event system [7] and use its subscribe-notify mechanism. The presence system of the IMS is specified in [8], the presence event package in [9].

The fundamental assumption that leads us to the reuse of the presence architecture is that we regard location information as a kind of presence information. The location of a user may be related to his presence state. The access to both presence and location data has the same requirements regarding security and privacy.

We propose the use of the Geography Markup Language (GML, [10]) to include location information in PIDF documents. Since both languages are XML based, GML elements can be integrated in Presence Information Data Format (PIDF) documents [11] easily.

Notification filters enable more complex notifications than pure location updates. The filter is sent within the subscribe message from the watcher to the presence server. For example the filter “(longitude TO centre BY distance) OR (latitude TO centre BY distance)” will send a notification whenever the presentity enters a rectangle around the centre coordinates. This reduces traffic over the air interface (single notification instead of constant location updates) at the price of added complexity at the presentity's terminal.

In order to access a terminal based location enabler, we propose to install a SIP Presence Edge Server at the GPS enabled user terminal. An edge presence server is a presence agent that is co-located with a Presence User Agent (PUA).

The simplest form of the proposed system architecture is sketched in figure 2: a peer-to-peer location presence system. The user's terminal hosts an edge presence server for location presence information. It allows 3rd party applications or other users to subscribe to location presence data notifications. A watcher application may use SIP messaging or interfaces like the ParlayX location API [12] to request presence location data. The client software in the user's terminal uses the J2ME Location API [13] to access the GPS module providing the physical location.

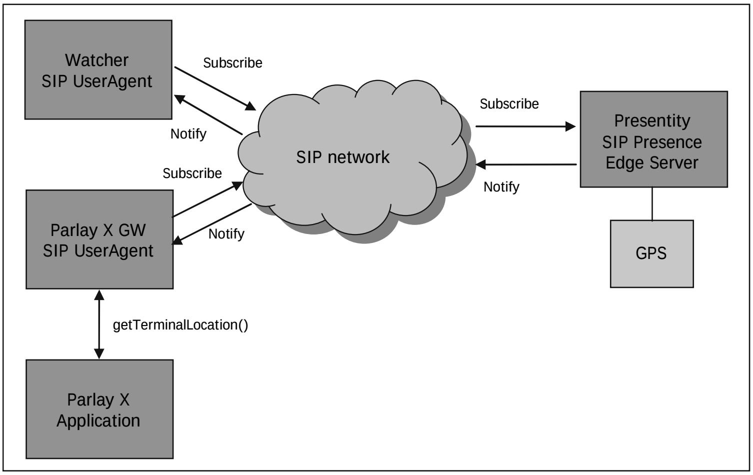


Fig. 2. Peer-to-Peer system Architecture

In order to support network-based location for terminals without a GPS receiver, we propose to use a Presence Agent in the network that aggregates presence data. The Presence Aggregator intercepts all subscription messages. Based on a user data base or on subscription policies, it decides how to locate the presentity. If the location is only available in the network, the subscription is forwarded to the presence server. If the user has an active GPS receiver, the subscription will be sent to the terminal. If both options are available, the aggregator can check the data for reliability. (Are the GPS coordinates within the cell?).

Location presence data show some essential differences from other presence data. First, location data is semantically different from classic presence data due to the almost continuous state space. Second, location data originates from different sources than normal presence state. As location data is calculated in the presentity's terminal, it requires a distributed approach. We therefore propose to separate subscriptions and notifications for location presence data from classic presence by introducing a new locpres: URI-scheme. The intended usage of the locpres: URI follows closely the usage of the pres: URI. However, it allows routing a subscription for presence location data to the presentity's terminal by addressing a subscription request to the locpres: URI of the target. In order to implement dedicated presence location access policies, filtering mechanisms and caching disciplines we propose publishing a locpres: URI of a presence location server that is dedicated to the handling of location data.

The bottleneck of today's mobile communication networks is the wireless access network – both in terms of delay and bandwidth. The communication between the mobile terminal and the location server uses the wireless access network, so traffic over this interface must be reduced to a minimum for optimum performance.

We argue that terminal-based filters can significantly reduce the amount of traffic on the air interface. The mobile terminal is not required to forward the location stream it receives from the GPS receiver to the LE.

The solution that we have chosen is to provide the mobile terminal with filter information within the location presence subscribe messages. Along with the description of the location presence event these messages contain an area and the action to be taken. E.g. the subscribe message requests the mobile terminal to generate a notification on entering or leaving a rectangle area specified by centre coordinates, length and width.

The sequence diagram in figure 3 shows the message sequence for a terminal-centric location service's operation.

The figure omits the watcher who subscribed for the specific service. E.g., user Alice asks for the service "notify me, if Bob is in his office". The presence service maps the logical location information "Bob's office" onto the corresponding geographical location data – e.g. a rectangle defined by its center coordinates, length and width. The presence service contacts the Location Enabler (LE) on behalf of the watcher Alice by sending a SIP subscribe message (1.0). Part of the subscribe message is the GML-encoded location that Alice is interested in and a description of the expected QoS: location accuracy, timing requirements, etc. The LE authorizes the subscribe request (1.1) and, if successful, forwards the subscribe message to the Edge Presence Server (EPS) on Bob's mobile device (1.3). Depending on the policy configured by Bob, the mobile device might ask interactively for Bob's explicit permission before accepting Alice's location request.

The GPS receiver sends continuous location updates to the mobile device, which are forwarded via the JSR179 interface [13] to the terminal's EPS (1.2, 1.4, and 1.5). The EPS detects on receipt of message 1.5 that Bob is located within the rectangle area the LE has subscribed to and that the QoS conditions are satisfied (accuracy is within limits). The EPS then sends a SIP notify message (1.7) to the LE. Depending on the QoS requested by Alice, the LE might decide to invoke a network-centric location mechanism (1.8) to do a consistency check (1.9) on Bob's location data. E.g., Alice might be a lawful authority representative who set the location request's QoS to the highest possible location reliability.

The LE then forwards the SIP notify message to the presence service (1.9), which processes the location data and then contacts the watcher Alice.

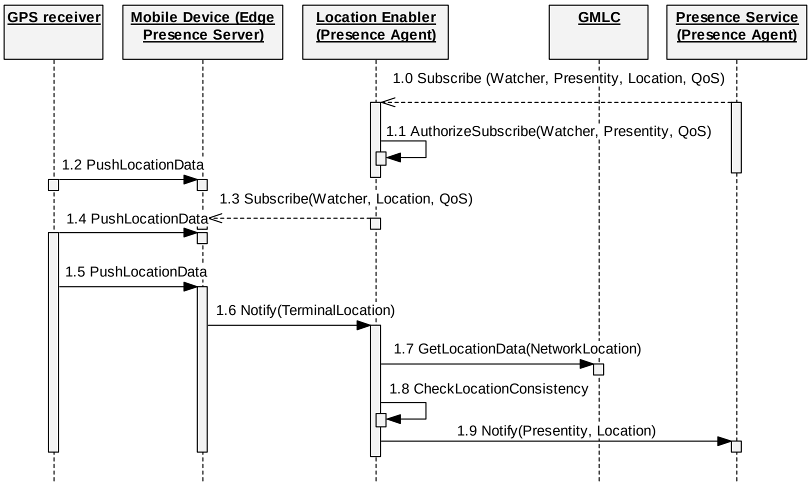


Fig. 3. Terminal-centric location enabler operation

5 IMS Presence Location Performance

A network-based positioning method requires request routing, access authorization and privacy control. Even if requester and target are connected to the same network, these functions are not trivial. In a roaming scenario, however, where a location request from a requesting network has to reach a location enabler in the visited network where the target terminal is actually roaming, is indeed a complex task, especially when access policies and privacy settings should be respected. The 3GPP Location Enabler architecture hence requires complex request routing between Requesting-, Home- and Visited network components in order to resolve the GMLC responsible for the user's terminal.

In an IMS-enabled system, however, the MSISDN based routing of a GMLC should be replaced by SIP message routing with its positive implications regarding performance, network management and security. Therefore our proposal is to send the location request directly to the user's URI using plain SIP message routing instead of introducing complex interworking functions in order to address a network-based location enabler.

Furthermore, the existing network-based positioning methods are either insufficient regarding accuracy or require up-front investments in the core network. Using a terminal-based GPS receiver as a source for location data is a network technology independent solution. It offers satisfactory accuracy while requiring only investments in those customers who want to use a location service.

Privacy requirements and mechanisms proposed by 3GPP standardization for Location Services are similar to those of presence systems [4] [14]. Beyond the specified functions, a real world location service system requires components that allow provisioning, user access and user control of privacy parameters. It seems reasonable to re-use the IMS presence infrastructure for subscription, authorization and privacy management of Location Based Services.

Finally, a network-based positioning method is forced to implement triggered location updates through polling. In order to conserve battery power mobile terminals tend to be in an idle state most of the time. Hence position changes are not visible to the network. We propose to implement trigger logic in the terminal so that necessary location updates are kept to an absolute minimum and scarce resources in the radio access network are used in the most economic way.

6 Related Work

The Geographical Location and Privacy (GEOPRIV) working group of the IETF [15] has investigated a number of problems related to the distribution of geographical information on the Internet, from both a security and a policy angle. The result of this work is a generic framework for the creation and distribution of location information on the Internet that enables confidentiality and policy directives, which are abstracted from the format of the location information.

Küpper and Treu propose complex location update strategies in the mobile terminal in order to realize scalable Location Based Services [16]. Shaham et al. propose to use SIP event mechanism to transport location data [17].

Polk and Rosen present a framework and requirements for usage of SIP to convey user location information and consider cases where message routing by intermediaries is influenced by the location of the session initiator [18].

7 Conclusion and Future Work

In this paper we have presented the architecture of a terminal-centric IMS location enabler. Our solution scales better and is more accurate, efficient and cost effective than current network-based location enablers.

Our group is currently working on an extension of SIP event filtering specifications and mechanisms in order to express and implement complex triggering criteria in the mobile terminal. Furthermore we are in the course of implementing a prototype system in order to prove our concept in a reference implementation.

Finally, we intend to contribute to the 3GPP/IMS standardization process with the aim to add the IMS location enabler to the family of basic enablers such as presence and messaging.

References

[1] G. Camarillo, M. A. Garcia-Martin, "The 3G IP Multimedia Subsystem (IMS)", John Wiley & Sons Ltd, 2004

[2] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; Location Services (LCS); Service description; Stage 1", 3GPP TS 22.071, version 7.1.0, 2005-01

[3] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol", IETF Standards Track RFC 3261, June 2002

[4] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; Location Services (LCS); Service description; Stage 1", 3GPP TS 22.071, version 7.1.0, 2005-01

[5] 3rd Generation Partnership Project, “Technical Specification Group Services and System Aspects; Functional stage 2 description of Location Services (LCS)”, 3GPP TS 23.271, version 7.0.0, 2005-03
[6] M. Day, J. Rosenberg, H. Sugano, ‘A Model for Presence and Instant Messaging’, IETF RFC 2778, February 2000
[7] A. B. Roach , “Session Initiation Protocol (SIP)-Specific Event Notification”, IETF Standards Track RFC 3265, June 2002
[8] 3rd Generation Partnership Project, “Presence Service; Architecture and Functional Description”, 3GPP TS 23.141, Version 6.7.0, 09/2004
[9] J. Rosenberg, ‘A Presence Event Package for the Session Initiation Protocol (SIP)’, IETF Standards Track RFC 3856, August 2004
[10] OpenGIS, "Open Geography Markup Language (GML) Implementation Specification", OGC 02-023r4, January 2003, http://www.opengis.org/techno/implementation.htm
[11] H. Sugano, S. Fujimoto, G. Klyne, A. Bateman, W. Carr, J. Peterson, “Presence Information Data Format (PIDF)”, IETF Standards Track RFC 3863, August 2004
[12] ETSI Standard ES 202 391-9, Open Service Access (OSA); Parlay X Web Services; Part 9: Terminal Location, V1.1.1 (2005-03)
[13] Java Specification Request JSR-179, ‘Location API for the Java 2 Platform, Micro Edition (J2ME) version 1.0’, Nokia, 2003. http://jcp.org/en/jsr/detail?id=179
[14] J. Peterson, “A Presence Architecture for the Distribution of GEOPRIV Location Objects”, IETF Internet-Draft, draft-ietf-geopriv-pres-02, September 8, 2004
[15] Charter of the Geographical Location and Privacy (GEOPRIV) working group of the IETF, http://www.ietf.org/html.charters/geopriv-charter.html
[16] A. Küpper, G. Treu: From Location to Position Management: User Tracking for Location-based Services. Kommunikation in Verteilten Systemen (KiVS) Kurzbeiträge und Workshop 2005, pp. 81-88
[17] R. Shaham, H. Schulzrinne, W. Kellerer, S.Thakolsri, An Architecture for LBS Mobility Using the SIP Event Model, International Conference on Mobile Systems Application and Services, (Mobys 2004) Boston 2004
[18] J. M. Polk, B. Rosen, “Requirements for Session Initiation Protocol Location Conveyance”, IETF Internet-Draft, draft-ietf-sipping-location-requirements-02.txt, October 25th, 2004
[19] R. Pailer, F. Wegscheider, S. Bessler, “A Terminal-Based Location Service Enabler for the IP Multimedia Subsystem”, IEEE Wireless Communications and Networking Conference, April 2006, Las Vegas, USA
[20] J. Fabini, R. Pailer, M. Happenhofer, “Terminal-Centric Location Services for the IP Multimedia Subsystem”, IEEE 63rd Vehicular Technology Conference, February 2006, Melbourne

WiKaF - A Knowledge-based Kalman-Filter for Pedestrian Positioning

Michael Thienelt, Andreas Eichhorn, Alexander Reiterer

Research Group Engineering Geodesy, Institute for Geodesy and Geophysics, Vienna University of Technology
andreas.eichhorn@tuwien.ac.at

Abstract

The precise, reliable and preferably ubiquitous positioning of mobile users is a substantial precondition for the provision of location and situation based information by Location Based Services. Within these applications, determining the location of pedestrians in 'passive environment' represents a special challenge.In this paper a new knowledge-based approach for the improvement of position quality is presented.

1 Introduction

Within the last years there was an enormous increase of various navigation applications. Vehicle navigation systems have makde the transition from the research laboratory to usable product. The next logical step is to navigate the individual person. Cellular phone network operators already have a major presence in the navigation market. This is motivated by the necessity to amortise their investments in expensive UMTS licenses introducing new profitable customer services, especially so called location based services (LBS). Many of these services offer location based information for users with mobile devices (e.g., PDA, mobile phone etc.). The service is performed depending on the current position/situation of the user. Typical examples are “Friendfinding” or monitoring of the closest restaurants or shopping opportunities.

The reorganization of the American army has also led to an intensified research work with LBS. One important emphasis is the positioning of pedestrians, in this case, the infantry. Navigation supported by mobile devices aims at a faster and less error-prone orientation in unknown areas.

Civilian applications are mainly focussed on tourist navigation. The individual pedestrian should be able to explore a city in an independent and comfortable way. Depending on his present position he gets additional information, for example of sights and shopping facilities.

Depending on the type of environment the methods for positioning (locating) can be divided in two groups: (1) locating in passive and (2) locating in active environments. Passive environments have no active sensor infrastructure. The locating of the user is realized autonomously by mobile devices with several sensors like GPS, mobile phone, digital compass or pedometer [1].

Active environments contain active landmarks. Active landmarks are single sensors or sensor networks which connect themselves with a mobile device transferring information about its present position. A successful application is realized in museums by infrared interfaces [2]. The exhibits locate the visitor and submit relevant information (e.g., historical data) to his mobile device. Using WLAN active landmarks can also be realized in urban areas. Current research activities try to integrate the RFID[1] technology into this development.

The common goal of all different positioning methods is to provide precise, reliable and preferably ubiquitous position information of the mobile users. It is a substantial precondition for the application of LBS. Especially the autonomous positioning of pedestrians in passive environments is influenced by many quality restrictions and consequently a special challenge. This contribution presents a new knowledge-based approach for the significant improvement of position quality: WiKaF (= ‘**Wi**ssensbasiertes **K**ALMAN-**F**ilter’, see [3]).

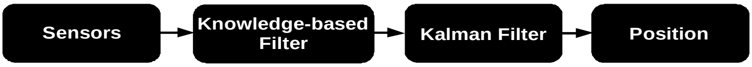


Fig. 1. Process chain of the WiKaF-System

[1] RFID (radio frequency identification) is a method for wireless reading and storing data. The data will be stored on special RFID tags. RFID contains the tags, a transmission/receiver unit and the integration of these components in an existing server system.

2 System architecture

In Figure 1 the process chain for the positioning of pedestrians with knowledge-based KALMAN-filtering is presented. A multi-sensor platform acquires absolute and relative position data of the pedestrian. In a knowledge-based pre-processing the measuring data is checked for plausibility and quality. Rough errors are eliminated and adequate stochastical models are selected/adapted dependent on the current environmental conditions of the pedestrian. In the next step a KALMAN-filter performs the combination of the hybrid sensor information (= multi-sensor integration) to an optimal estimation of the kinematic state of the pedestrian (e.g. position, speed and direction of the movement). In the following section the individual components of the process chain are described.

2.1 Sensors

WiKaF uses the following sensors: Vaisala PTB 220, Honeywell HMR 3000, Garmin eTrex Summit and a Dead Reckoning module (DRM III) from the company PointResearch. All sensors have been acquired in the NAVIO project [4] and are presented in Figure 2.

The **PTB220** is a fully compensated digital barometer and covers a wide measuring range of air pressure: 500 to 1100 hPa. It has a special silicon-sensor that was developed by the company Vaisala for precise barometric measurements. The sensor is non-sensitive to changes in temperature (range of application: -40 to +60°C), vibrations and shows a good long term stability [5]. In WiKaF it shall be used for indoor navigation (detection of different floors in buildings).

The **HMR 3000** is a compass module which consists of three magnetoresistive sensors and a two-axis tilt sensor to produce tilt compensated heading data [6]. The determination of the heading is an important precondition for positioning in a dead reckoning scenario.

The **eTrex Summit** is a 12-channel handheld GPS receiver [7]. It consists of an integrated GPS antenna, a digital compass and a barometric altimeter. It runs up to 22 hours with batteries (save mode) and was chosen to get redundant GPS data (see also DRM III).

The **DRM III** is a commercial sensor module for pedestrian navigation [8]. It combines a data acquisition (absolute and relative sensors) with a processing component (KALMAN-filter). In WiKaF the module is only used for data acquisition. Data processing will be realized with the new knowledge-based KALMAN-filter. Nevertheless the trajectory processed by the DRM III can be used for comparison and evaluation of the results.

The module consists of a 12 channel GPS receiver with antenna, a digital compass, a triax accelerometer (usable for step recognition), a barometer and a gyro. The GPS antenna can be fastened on the head of the pedestrian. The azimuth of the movement is determined by a combination of digital compass and RateGyro. Number of steps and a scale factor (step length) enable to derive the covered distance. For step recognition it is very important that the module is centrically fastened at the belt (back, see Figure 2) of the user. The correct adaptation of the module has a significant influence on the quality of step recognition.

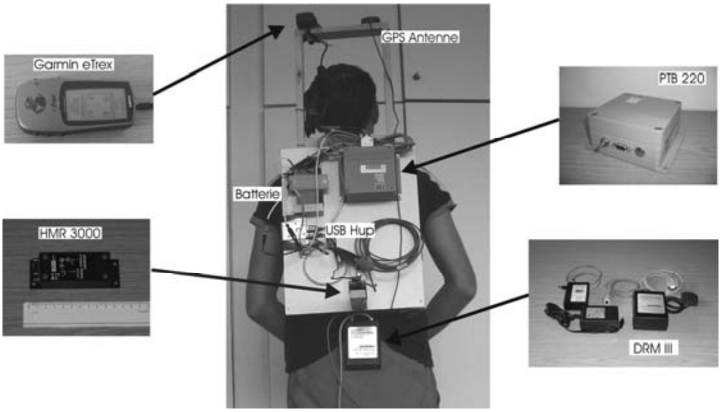


Fig. 2. Integration of the sensors into a multi-sensor system

2.2 Knowledge-based filtering

After the data acquisition by the multi-sensor system the measuring quantities are submitted to a first filter process. The task is to detect disturbances and outliers (which shall not influence the KALMAN-filter), to eliminate them and to create specific error models for the measuring data adapted to the current environmental situation. For successful filtering extensive ‘knowledge engineering’ has to be done within a recursive process (see Figure 6). Therefore a complete decoupling of this system component is envisaged. On basis of this requirement the filtering is realized as a knowledge-based component. The separation of problem knowledge and knowledge processing enables a more easily extension, modification or exchange of the knowledge base.

The advantages of a knowledge-based approach in comparison with a conventional one (realized in Delphi, Fortran or C++) are: (1) the knowledge about the problem domain is separated from general problem-solving knowledge (makes it easier for the knowledge engineer to manipulate this knowledge); (2) experts-knowledge, exists very often in form of rules, can be captured in this form without converting into forests of data definitions and procedures.

Independent from the point of view, knowledge-based systems consist of the following major components: a knowledge base, an inference engine, a user interface, a knowledge acquisition tool and an explanation tool [9]. The knowledge base, the most important component of a knowledge-based system, contains the relevant domain-knowledge that the knowledge engineer has implemented in the course of the development process. The process of deriving a solution for a problem can be viewed simplistically as one of finding a connection between the input and the conclusion; this process is accomplished by the inference engine. The user interface (UI) serves to provide the end user and the knowledge engineer with a friendly means of communication with the program system. The knowledge engineer has to interact with the domain specialist to acquire knowledge about the problem domain and to implement this knowledge in the knowledge base. The knowledge acquisition tool assists the knowledge engineer in this work. The explanation facility helps the user to understand the knowledge reasoning of the system. Explanations (automatically or if required) enhance the usefulness and acceptance of knowledge-based systems.

Several declarative knowledge representation schemas are commonly used: predicate logic, production rules, semantic nets, frames and others. Our knowledge base is realised as a rule based/object-oriented approach. Rule-based programming is one of the most commonly used techniques for implementation of a knowledge base. Rules are used to represent heuristics, or ‘rules of thumb’, which specify a set of relationships. A rule-based approach consists of two parts: a set of rules and a working memory. A rule is divided into two parts, namely the left hand side (LHS) and the right hand side (RHS). In the LHS, we formulate the preconditions of the rule, whereas in the RHS the actions are formulated. A rule can be applied (or fired), if all its preconditions are satisfied; the actions specified in the RHS are then executed. The second component of a rule-based system, the working memory, is a collection of working memory elements, which itself are instantiations of a working memory type (WMT). WMTs can be considered as record declarations in PASCAL or strict declarations in C. There are algorithms for the matching phase, i.e., the phase where all rules are checked against all working memory elements, which are efficient in practice. The result of the matching phase is the conflict set, that includes all rule instances 'ready to be fired'. A conflict resolution strategy selects one rule instance which is actually fired.

2.3 Central KALMAN-filter for optimal multi-sensor integration

The purpose of the multi-sensor integration is to merge the (knowledge-based) pre-processed measuring data to accurate, reliable and preferably everywhere available position information. The fusion shall be realized by the creation of a 'Position Finding Module (PFM)' with a central KALMAN-filter. Using theoretical assumptions concerning the motion of the pedestrian in one scanning interval (system equations) and empirical measurements (measuring equations), KALMAN-filtering enables the optimal estimation of the state of motion of the pedestrian in each epoch (see Figure 3, left).

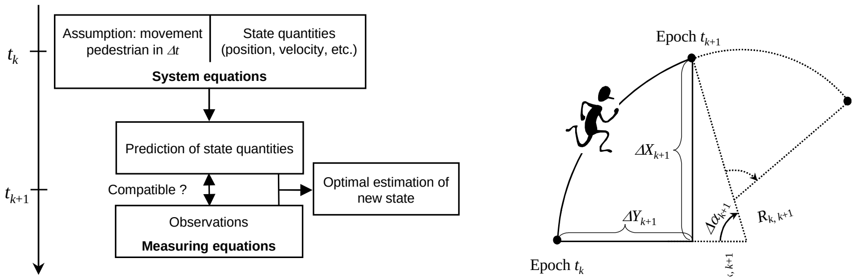


Fig. 3. KALMAN-filter and constant circular motion (2D)

Between two filter epochs t_k and t_{k+1} the (non-disturbed) motion of the pedestrian is assumed to be constant circular. For 2D-coordinates a possible analytical model is a series of circular arcs with common tangents (see Figure 3, right). An integrated 3D-model (helix) is used in [10]. In case of sudden changes of orientation of the pedestrian kinematic models normally show a certain inertia. This effect can be significantly reduced by integration of additionally measured correcting variables into the system equations. The 'causative modification' of kinematic models is successfully applied in [11].

In Figure 4 the architecture of the PFM is shown. It contains 4 sub-modules which consider the environment ('Outdoor' and 'Indoor'), the kinematic state of the pedestrian ('Pedestrian stop') and emergency situations ('Failure').

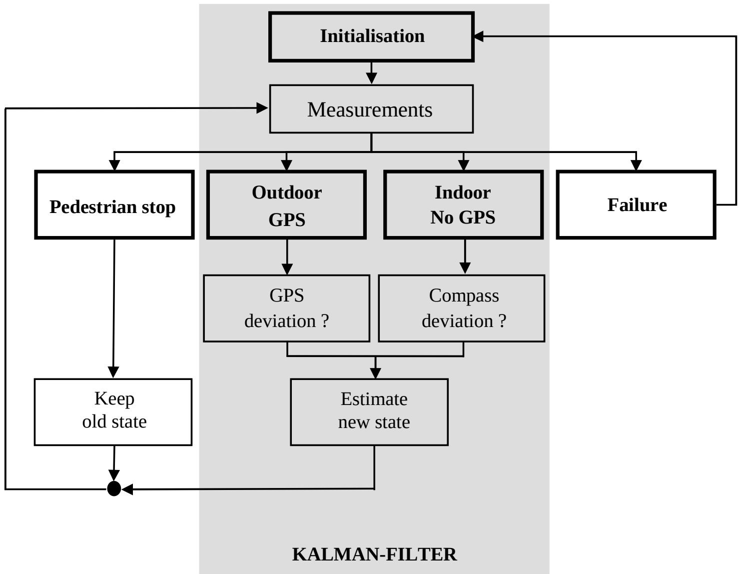


Fig. 4. Architecture of the PFM

The detection of bad GPS quality (multipath effects etc.) in outdoor situations is a special challenge for the PFM. Using the results from the knowledge-based pre-processing, the KALMAN-filter enables an additional statistical analysis of the contradictions between motion model and GPS-observations (i.e. innovation-tests).

In geodetic applications the stochastically models of measuring and disturbance quantities are usually assumed to be white noise processes with no correlations in time. In the case of high measuring frequencies there often appear high autocorrelations (coloured noise) between succeeding observations. Using a 'form-filter' extension, these correlations can be integrated into the classical KALMAN-filter algorithm. The central problem is the determination of the correct parameters of the noise process. One goal of WiKaF is to investigate the possibilities and the efficiency of the quantification of autocorrelation functions with different strategies:

1. Determination of the relevant parameters of the autocorrelation functions by a priori time-series analysis
2. Estimation of the parameters during the filter process using an adaptive extension (adaptive KALMAN-filtering)

3 Further workflow of WiKaF

Having shown the theoretical fundamentals and the system architecture, the workflow for the realization of WiKaF is pointed out. It includes the definition of suitable Test areas and the detailed planning of the individual levels of development.

3.1 Test areas

The development and implementation of the knowledge-based KALMAN-filter requires the definition of different types of representative Test areas. In the beginning of the project three scenarios were selected: two outdoor scenarios and one indoor scenario.

Test area 1 (TA-1) is realized at the park of the castle Schoenbrunn in Vienna and offers excellent GPS-quality. Test area 2 (TA-2) is next to the centre of the city and stretches a triangle of 140m x 200m x 250m. This area is located in an urban environment with partly poor or no GPS availability. 3D-reference trajectories are realized by tachometry and precise levelling.

Figure 1 in chapter 15 shows the 3D-reference trajectory for the indoor scenario (Test area 3 (TA-3)) where the trajectory is realized from the ground to the third floor of the office building of the Vienna University of Technology. At the half distance there is a little ‘side trip’ to a room which is accessible from the corridor. The endings of TA-3 are directly connected with TA-2 and enable the change from an outdoor to an indoor scenario.

3.2 Workflow of WiKaF

At the end of the conceptual description the further project steps will be briefly presented. A rough overview of the most important project phases is shown in Figure 5.

The selection of the different types of sensors (see also Section 2.1) and the creation of representative reference trajectories is completed. The sensor investigation and the derivation of deterministic and stochastic sensor and disturbance models is just in work. As main part of the ‘knowledge engineering’ these investigations represent one of the most important steps within the WiKaF project. The collected knowledge will be directly integrated into the two following development steps.

The parallel and in most parts independent development of knowledge-based system and KALMAN-filter (PFM) leads to functionally autonomous system components. It enables the establishment of a modular system and a more simple evaluation of the knowledge-based approach against ‘classical’ KALMAN-filtering by activating/deactivating of one system component.

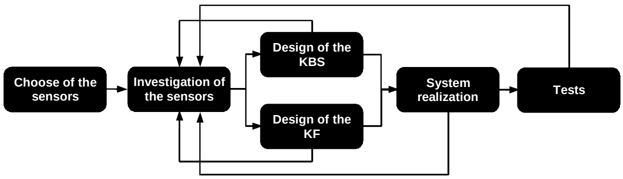


Fig. 5. Workflow of WiKaF

After finishing the conceptual design the complete positioning system will be realized by two different software solutions. The implementation of the knowledge-based system component will be carried out by Clips respectively WxClips. Clips is a programming language especially designed for the creation of rule-based expert systems (see [12]). The KALMAN-filter and the PFM will be realized by Matlab. Data exchange (interfaces) between the different software modules happens by standardized ASCII-files.

4 Conclusions and perspectives

The concept of knowledge-based KALMAN-filtering (WiKaF) is an ‘intelligent’ extension of the data-based multi-sensor integration. It enables the utilization of not only quantitative (measurements) but also qualitative information by the definition of rules. The major goal is to support the filter in detecting and eliminating of incorrect sensor data and selecting suitable (situation related) stochastical models. In this contribution we report on the architecture and the functionality of WiKaF and its present stage of development.

Using the reference trajectories described in Section 3.1, the next part of the project deals with the investigation of the single sensor systems. In this context the time series analysis of the sensor data, the sensor calibration and the evaluation and classification of topographical (disturbance) influences are important tasks and represent the basis for the design of the knowledge-based component. A special emphasis is set on the determination of auto-correlation functions as input for the form-filter (see Section 2.3).

Acknowledgements

The research work presented in this paper is supported in parts by the FWF Project NAVIO of the Austrian Science Fund (Fonds zur Förderung wissenschaftlicher Forschung) (Project No. P16277-N04)

References

[1] Retscher, G. and Thienelt, M.: NAVIO - A Navigation and Guidance Service for Pedestrians. Journal of Global Positioning Systems (Print Version), 3 (2004),1-2; pp. 208-217
[2] Eyeled-Website: http://www.eyeled.de, last access: 09.2005
[3] Thienelt, M. / Eichhorn, A. / Reiterer, A.: Konzept eines wissensbasierten Kalman-Filters für die Fußgängerortung (WiKaF), Österreichische Zeitschrift für Vermessung und Geoinformation (VGI), 93. Jahrgang, Heft 2/2005, pp. 96-104
[4] Gartner, G./ Frank, A. / Retscher, G.: Pedestrian Navigation System in Mixed Indoor/Outdoor Environment – The NAVIO Project, in: Schrenk M. (Hrsg.): Proceedings of the CORP 2004 and Geomultimedia04 Symposium, February 24-27, 2004, Vienna, Austria, 165-171
[5] Vaisala–Website: http://www.vaisala.com/businessareas/instruments/products barometric pressure/ptb220, last access 09.2005
[6] Honeywell-Website: http://www.ssec.honeywell.com/magnetic/datasheets/ hmr3000.pdf, last access 09.2005
[7] Garmin-Website: http://www.garmin.de/Produktbeschreibungen/GPSetrex.pdf, last access 09.2005
[8] PointResearch: http://www.pointresearch.com/drm_eval.htm, last access 09.2005
[9] Stefik, M.: Introduction to Knowledge Systems. 2. Aufl., Morgan Kaufmann Verlag, San Francisco, 1998
[10] Aussems, T.: Positionsschätzung von Landfahrzeugen mittels Kalman-Filterung aus Satelliten- und Koppelnavigationsbeobachtungen. Veröffentlichungen des Geodätischen Instituts der Rheinisch-Westfälischen Technischen Hochschule Aachen, Nr. 55, Aachen, 1999
[11] Eichhorn, A.: Ein Beitrag zur Identifikation von dynamischen Strukturmodellen mit Methoden der adaptiven Kalman-Filterung. DGK, Reihe C, Nr. 585, München, 2005
[12] Clips-Website: http://www.ghg.net/clips/CLIPS.html, last access: 09.2005

Map-independent positioning of land vehicles with causative modified motion equations

Andreas Eichhorn

Institute for Geodesy and Geophysics, Engineering Geodesy, Vienna University of Technology
andreas.eichhorn@tuwien.ac.at

Abstract

In this paper some results of the project 'Positioning Component' are presented, which was subject of a cooperation between the University of Stuttgart and DaimlerChrysler AG. The aim of the project was to develop a map-independent module for the autonomous positioning of vehicles within the framework of Location Based Services (LBS). The integration of the absolute position information (GPS) and the relative information (differential odometer and gyro) is realized with a central KALMAN-filter. The system equations of this filter originally base on a kinematic model (constant velocity in one sampling interval). To reduce the filter's inertia in the case of fast cornering it is necessary to modify these equations in a causative sense: the measured yawing (derived from the gyro) is used as a 'geometrical' correcting variable. The presented examples of filter results show different representative driving scenarios in inner urban areas with bad GPS quality and on highways. The mean errors of the estimated positions vary from approx. s_P = 1,9 m (on highways) to 3,0 m (in urban areas with dense buildings and tunnels). The accuracy and availability requirements of the car manufacturer are achieved.

1 Motivation

In the future many applications linked with Location Based Services (LBS) will need an accurate and in particular a reliable positioning of the vehicle at any moment. A typical application is the automatic emergency call in case of a traffic accident. The additional transmission of the last sequence of georeferenced vehicle positions to an emergency call center enables the guidance of rescue teams to the scene of accident (see Figure 1).

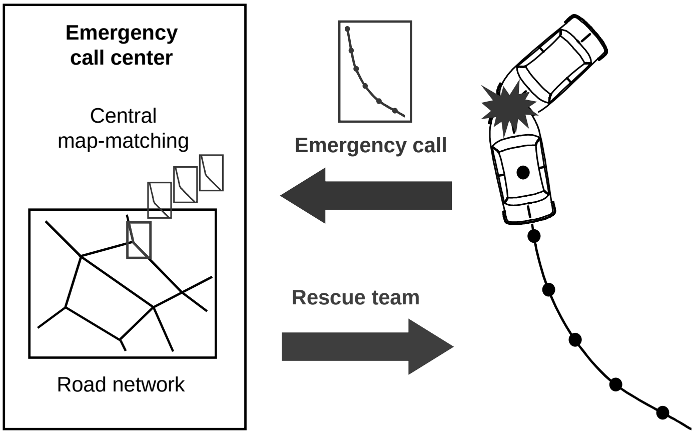


Fig. 1. Automatic emergency call with central map-matching process

In contrast to common car navigation systems this service does not need a digital map onboard. The navigation is realized by a map-matching process which is directly executed in the emergency call center itself. The central storage of road data is more cost effective and easier to keep up-to-date. Consequently the bord-autonomous part of this process is map-independent.

The project 'Positioning Component' (see [1], [2] and [3]) was subject of a cooperation between the University of Stuttgart and Daimler Chrysler AG. The aim of the project was to develop a map-independent module for the autonomous positioning of vehicles. The main requirements were defined as:

- Real-time capability of the algorithm, this means calculation of the vehicles position with a time delay of max. some msec.
- No highly accurate ($\sigma Y = \sigma X < 4$ m) but track stable position results

- Good geometrical representation of the passed trajectory (especially turns) as precondition for the central map-matching process realized by a sample rate of 1 Hz
- Reliable and always available position information using lowcost sensors from the vehicles standard equipment (i.e., GPS, differential odometer from the anti-lock braking system and gyro from the stabilization system)
- Application in different driving scenarios, this means public traffic on highways and in cities with partly very bad GPS-quality respectively no GPS available.

The creation of a ‘Position Finding Module’ integrating a multi-sensor system in a central KALMAN-filter was decided to be the best strategy for fulfilling all requirements.

2 Multi-sensor system

As shown in Figure 2 the chosen sensor configuration consists of GPS, differential odometer (from anti-lock braking system) and a gyro (from the stabilization system).

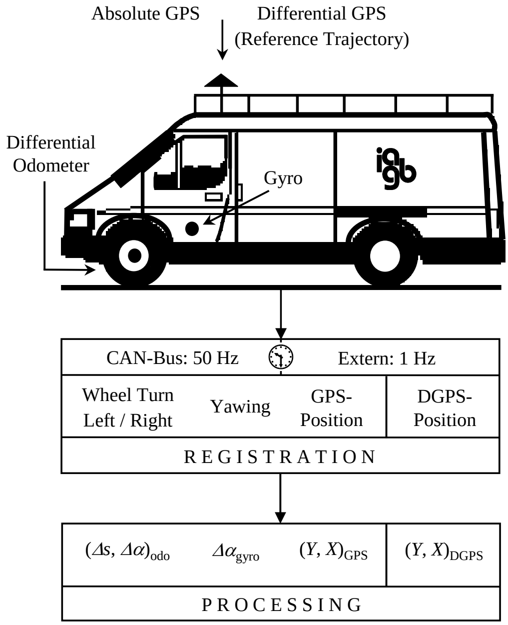


Fig. 2. Sensor configuration and derivation of geometrical quantities

The sensors are representing a combination of absolute (GPS) and relative position information (differential odometer and gyro). In the case of no GPS available (e.g. in tunnels) it is possible to calculate the vehicles position in dead reckoning. Redundant position information enables mutual control and error detection. Using this combination, reliability as well as availability of the individual sensor is increased.

During the test-drives the registration of differential odometer and gyro was realized via the vehicles CAN-bus (CAN = Controller Area Network) with a frequency of 50 Hz. Absolute GPS and differential GPS (DGPS) were separately registered with a frequency of 1 Hz. DGPS was not used for position finding, but to derive the reference trajectories for quality control of the results. The synchronization of all signals was better than 2 ms. Assuming a velocity of 200 km/h, this means a position error (longitudinal deviation) of only 10 cm [4].

The second step contains the derivation of geometrical quantities from the original signals. The wheel turn impulses are integrated from 50 Hz to 1 Hz and multiplied with a scale factor (from calibration) to calculate the distance covered and the change in the vehicles direction. The integration of the gyro signal delivers the change in direction as redundant information. A detailed description of the formulas can be found in [5]. The resulting position information (per s) is:

$$\boldsymbol{L}(t_{k+1}) = \boldsymbol{L}_{k+1} = \left(\begin{array}{cc} Y_{\mathrm{DGPS}} & X_{\mathrm{DGPS}} \\ \hline Y_{\mathrm{GPS}} & X_{\mathrm{GPS}} \\ \Delta s_{\mathrm{odo}} & \Delta\alpha_{\mathrm{odo}} \\ \Delta\alpha_{\mathrm{gyro}} & \end{array} \right)_{k+1} \qquad (1)$$

Sensor examination and calibration was realized by test-drives in three different scenarios: inner urban areas, highways with high velocity and winding country roads. The empirical standard deviations of the geometrical quantities (1) are shown in Table 1.

Table 1. Stochastical model of the multi-sensor system

	velocity [m/s]		
	0-5	5-15	> 15
$(s_Y, s_X)_{\mathrm{GPS}}$	1,5 m		
$s_{\Delta s,\ \mathrm{odo}}$ per s	0,2 m		0,3 m
$s_{\Delta\alpha,\ \mathrm{odo}}$ per s	2 gon		1 gon
$s_{\Delta\alpha,\ \mathrm{gyro}}$ per s	2 gon	1 gon	

3 KALMAN-filter algorithms with causative motion equations

To bring together all different sensor information under additional consideration of its individual stochastical features, a central KALMAN-filter (see e.g. [6] or [7]) is used. The recurrent construction of the algorithm enables at each time t_k the optimal estimation of vehicles position and other kinematic parameters (e.g. velocitiy). The KALMAN-filter is a suitable tool for real-time position finding and error detection.

3.1 Fundamentals

The KALMAN-Filter consists of two parts:

1. The theoretical system equations contain geometrical and physical assumptions concerning the motion of the vehicle in one scanning interval Δt (= 1 s).
2. The measuring equations represent the empirical part containing the observations derived from the sensors.

The combination of prediction (from system equations) and measurements enables to calculate the optimal estimation of the state vector, which contains the unknown kinematic parameters: vehicles position $(Y, X)_{k+1}$, orientation α_{k+1} and velocity v_{k+1}. One important feature of the algorithm is the possibility to test the compatibility of predicted and measured motion. Consequently the test provides the basic information for the detection of bad GPS-quality in inner urban areas.

3.2 Causative modified motion equations

The geometrical and physical fundamentals of the kinematic model are based on [8]. In one scanning interval Δt the vehicle is assumed to perform a constant circular motion (see Figure 3). This means constant velocity v, angular velocity $\dot{\alpha}$ and curvature $1/R$ in (t_k, t_{k+1}).

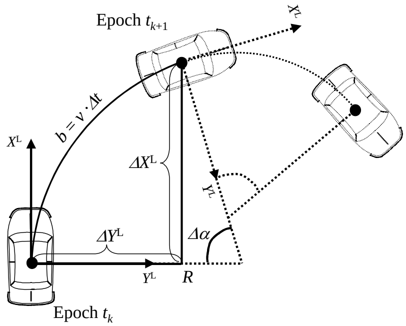


Fig. 3. Constant circurlar motion

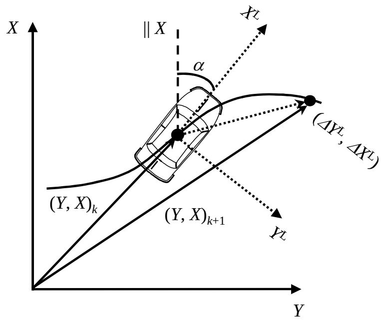


Fig. 4. Transformation into ref. frame

These assumptions are motivated by observations of vehicles kinematics. Possible accelerations are considered as disturbances *w*. The derivation of the deterministic system equations is realized in two steps. In the first step the motion is quantified in a local co-ordinate system (Figure 3) which in the second step is transformed into the global reference frame (Figure 4). The resulting recurrent non-linear system equations are shown in (2).

$$
\begin{aligned}
\begin{pmatrix} Y_{k+1} \\ X_{k+1} \end{pmatrix} &= \begin{pmatrix} Y_k \\ X_k \end{pmatrix} + \frac{v_k\, \Delta t}{\Delta\alpha_{k+1}} \begin{pmatrix} \cos(\alpha_k) & \sin(\alpha_k) \\ -\sin(\alpha_k) & \cos(\alpha_k) \end{pmatrix} \begin{pmatrix} 1-\cos(\Delta\alpha_{k+1}) \\ \sin(\Delta\alpha_{k+1}) \end{pmatrix} \\
\alpha_{k+1} &= \alpha_k + \Delta\alpha_{k+1} \\
v_{k+1} &= v_k
\end{aligned}
\tag{2}
$$

The state vector $\boldsymbol{x}$ consists of four elements: co-ordinates (Y, X), orientation (azimuth) α and velocity v of the vehicle. In difference to [8] the filter progress is realized using the change in direction $\Delta\alpha_{k+1}$ as correcting variable u, derived by gyro observations (see (1)). The main advantage of this 'causative modification' is the reduction of the filters inertia in the case of fast cornering. Extreme small radiuses are passed without any oscillations in the estimated trajectory. Representing the non-linear equations (2) by a vector function $\boldsymbol{\Psi}$, the causative modification is illustrated in (3) and (4):

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Psi}(\boldsymbol{x}_k, u_k = \Delta\alpha_{\text{gyro},k+1}) \qquad \text{non linear model} \tag{3}$$

$$\underset{4,1}{\underline{\boldsymbol{x}_{k+1}}} = \underset{4,4}{\underline{\boldsymbol{T}_{k+1,k}}}\,\underset{4,1}{\underline{\boldsymbol{x}_k}} + \underset{4,1}{\underline{\boldsymbol{B}_{k+1,k}}}\,\underset{1,1}{\underline{u_k}} + \underset{4,1}{\underline{\boldsymbol{S}_{k+1,k}}}\,\underset{1,1}{\underline{w_k}} \qquad \text{linear model} \tag{4}$$

In (4) additional disturbance accelerations (w = 1 m/s^2) are added. The JACOBI-matrices are indicated with $\boldsymbol{T}$ = transition matrix, $\boldsymbol{B}$ = coefficient matrix of correcting variable and $\boldsymbol{S}$ = coefficient matrix of disturbance variable.

The measuring equations are derived using the remaining position information of absolute GPS and differential odometer.

$$\begin{aligned} Y_{\text{GPS},k+1} - \varepsilon_Y &= Y_{k+1} \\ X_{\text{GPS},k+1} - \varepsilon_X &= X_{k+1} \\ \Delta s_{\text{odo},k+1} - \varepsilon_{\Delta s} &= v_{k+1}\,\Delta t \\ \Delta\alpha_{\text{odo},k+1} - \varepsilon_{\Delta\alpha} &= \alpha_{k+1} - \alpha_k \end{aligned} \tag{5}$$

4 Creation of a 'Position Finding Module'

To accomplish the requirements defined in section 1 and as precondition for operational use the KALMAN-filter is embedded in a 'Position Finding Module' (= PFM). The architecture of the module is shown in Figure 5.

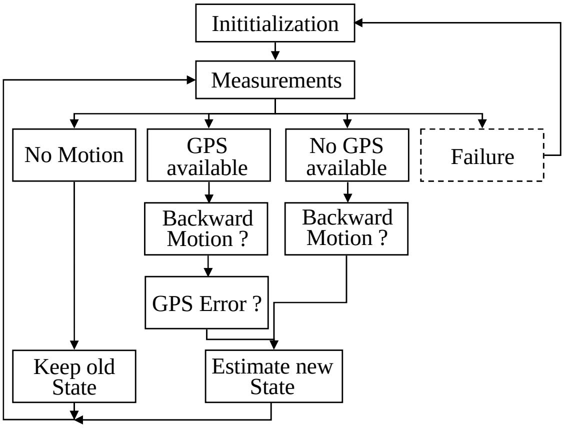


Fig. 5. Architecture of the PFM

The main task of the PFM is to create highly available and reliable position information which is directly linked with a stable filter progress. The module consists of four sub modules containing alternative strategies for position finding. The sub modules 'GPS available' and 'No GPS available' represent the central elements of PFM and will be shortly characterized.

In the case of GPS available (existing observation at filter input), several testing strategies are effected to judge its quality. The basic idea is to calculate the deviations between predicted (2) and observed vehicle position (5).

$$\boldsymbol{d}_{k+1} = \begin{pmatrix} d_{Y_{\mathrm{GPS}},\,k+1} \\ d_{X_{\mathrm{GPS}},\,k+1} \end{pmatrix} = \begin{pmatrix} Y_{\mathrm{GPS},\,k+1} - Y_{k+1} \\ X_{\mathrm{GPS},\,k+1} - X_{k+1} \end{pmatrix} \tag{6}$$

In inner urban areas multi path-effects and bad satellite geometry often cause large errors in GPS-position (> 10 m). Using the co-ordinate innovations (6), it's very simple to detect single outliers. A more pretentious task is the detection of parallel shifted trajectories which show nearly no difference in inner geometry but may induce a position error of more than 20 m. In PFM the identification is realized checking the profile of succeeding co-ordinate innovations. In case of a parallel trajectory the innovations are not randomly distributed but possess a systematic progress (see Figure 6, right). Replacing the absolute GPS-observations by GPS-differences

$$\begin{pmatrix} \Delta_{Y_{\mathrm{GPS}},\,k+1} \\ \Delta_{X_{\mathrm{GPS}},\,k+1} \end{pmatrix} = \begin{pmatrix} Y_{\mathrm{GPS},\,k+1} - Y_{\mathrm{GPS},\,k} \\ X_{\mathrm{GPS},\,k+1} - X_{\mathrm{GPS},\,k} \end{pmatrix} \tag{7}$$

the maintenance of inner geometry is used in dead reckoning to support differential odometer and gyro. In Figure 6 (left) a successfully detected parallel GPS-trajectory with a length of more than 300 m and its correction in PFM are shown.

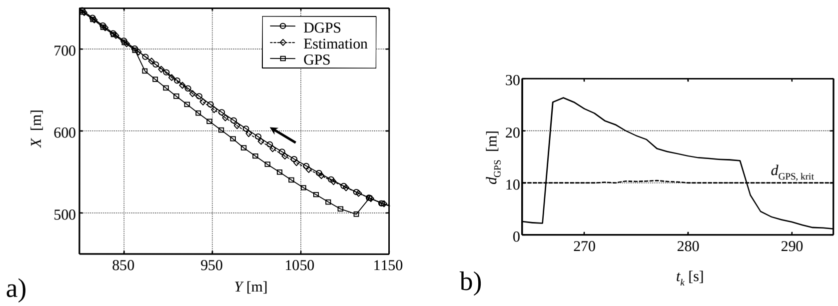


Fig. 6. Detection of systematic GPS-deviations

In the case of no GPS available (e.g. in tunnels) the accurate position is estimated in dead reckoning using only the causative modified motion equations and relative position information derived from differential odometer and gyro.

5 Results from different test-drives

By showing sections of representative test-drives some results are presented. In Figure 7 a typical scenario from inner urban area is presented. Shortly after cornering GPS-quality (square) breaks down and induces maximal position errors of more than 20 m. Consequently PFM has to estimate (diamond) the reference-trajectory (circle) reducing GPS-weights and increasing the influence of differential odometer and gyro. Using this relative position information and the inertia reduced causative motion equations, the geometry of the turn is very well represented by estimation. Transforming the filter results in a digital map, it can be shown that the deviations with measured reference trajectory are caused by a systematic drift in DGPS.

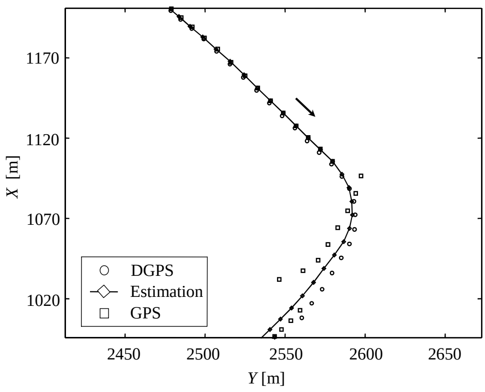


Fig. 7. Inner urban area: 90 degree cornering with bad GPS

As part of a highway-drive PFM passes in Figure 8 a 500 m tunnel in dead reckoning, this means no GPS available for 83 s and 44 s on both ways. The filter results never lose their relationship to the road, which demonstrates the rise of availability using PFM. At the beginning / end of the tunnel the cornerings represent very well the geometry of the turns.

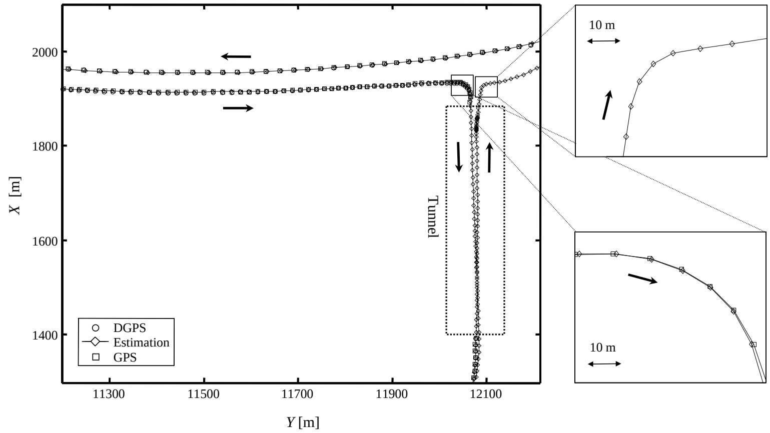


Fig. 8. Passing a 500 m tunnel to a highway in dead reckoning

The quality of the results is evaluated calculating empirical accuracy-measures derived from estimated trajectory and DGPS-position (high available differential-code-solution), e.g. s_{abs} (= standard deviation calculated with residuals from estimation ⇔ DGPS representing the absolute accuracy of position finding). Some results derived from several test-drive scenarios are presented in Table 2.

Table 2. Quality results from test-drives

Scenario	Length [km]	Filter epochs	Good GPS quality [%]	s_{abs} [m]
Inner urban areas	12,8	2097	76	2,85
High-density area	7,1	1425	72	3,04
Low-density area	5,7	672	85	2,49
Highways	27,5	1402	87	1,85

The mean absolute accuracy of PFM can be specified within a range of $s_{abs} \approx 1{,}9$ m (highways with good GPS quality) to $s_{abs} \approx 3{,}0$ m (inner urban areas with dense buildings and decreasing GPS quality). There are no significant differences between longitudinal and transverse deviations. Consequently the correct weighting of the different sensors can be assumed.

6 Conclusions

The research has shown that the requirements defined at the beginning could be achieved by using PFM. The accuracy is significantly better than the boundaries defined with $\sigma_Y = \sigma_X < 4$ m. It is also shown that the causative modification of the motion equations reduces the filter's inertia and enables a good geometrical representation of turns.

Bad results were obtained in the case of dead reckoning with wrong initial orientation of the vehicle (e.g., some test-drives through tunnels). In this case the information derived from motion equations and remaining relative position sensors is not sufficient to detect a rotation in the estimated trajectory which at its end may induce a deviation of several hundreds of meters. This effect defines the PFM's limits. It can only be compensated using additional absolute sensors (e.g. a compass) or information from a digital map.

References

[1] Eichhorn, A. / Kuhlmann H. / Bettermann, R. / Czommer, R.: Entwurf eines Kalman-Filters zur kartenunabhängigen Fahrzeugortung. Final project report, Stuttgart, 2001
[2] Eichhorn, A.: Ein Beitrag zur Identifikation von dynamischen Strukturmodellen mit Methoden der adaptiven Kalman-Filterung. DGK, Reihe C, Heft Nr. 585, München, 2005
[3] Eichhorn, A.: Development of a Kalman-filter module for real-time and map-independent positioning of land vehicles based on causative motion equations. VehCom-2003: Sensors, Navigation and Communications for Vehicles Telematics, Birmingham, 2003
[4] Kuhlmann, H. / Eichhorn, A.: Positionsbestimmung von PKW mit Sensoren der Serienausstattung. DGON-Symposium Posit. und Navigation, Dresden, 2003
[5] Czommer, R.: Leistungsfähigkeit fahrzeugautonomer Ortungsverfahren auf der Basis von Map-Matching-Techniken. DGK, Reihe C, Heft 535, München, 2001
[6] Brown, R.G. / Hwang, P.Y.C.: Introduction to Random Signals and Applied Kalman Filtering. John Wiley & Sons, New York, 1997
[7] Gelb, A.: Applied optimal Estimation. M.I.T. Press, Cambridge, Mass., 1974
[8] Aussems, T.: Positionsschätzung von Landfahrzeugen mittels Kalman-Filterung aus Satelliten- und Koppelnavigationsbeobachtungen. RWTH Aachen, Nr. 55, 1999

Section III Modelling and Awareness

Chapter 13	Steffen Volz Shortest Path Search in Multi-Representation Street Databases
Chapter 14	Inessa Seifert, Thomas Barkowsky, Christian Freksa Region-Based Representation for Assistance with Spatio-Temporal Planning in Unfamiliar Environments
Chapter 15	Mohammad Malek, Andrew Frank, Mahmoud Delavar A Logic-Based Foundation for Spatial Relationships in Mobile GIS Environment
Chapter 16	Kai-Florian Richter From Turn-By-Turn Directions to Overview Information on the Way to Take
Chapter 17	Masatoshi Arikawa, Kouzou Noaki Geocoding Japanese Walking Directions using Sidewalk Network Databases
Chapter 18	Tumasch Reichenbacher The concept of relevance in mobile maps
Chapter 19	L. Tiina Sarjakoski, Tommi Koivula, Tapani Sarjakoski A Knowledge-Based Map Adaptation Approach for Mobile Map Services
Chapter 20	Albrecht Weiser, Alexander Zipf A visual editor for OGC SLD files for automating the configuration of WMS and mobile map applications
Chapter 21	Rainer Simon, Harald Kunczier, Hermann Anegg Towards Orientation-Aware Location Based Mobile Services
Chapter 22	Patrick Luley, Lucas Paletta, Alexander Almer, Mathias Schardt, Josef Ringert Geo-Services and Computer Vision for Object Awareness in Mobile System Applications

Shortest Path Search in Multi-Representation Street Databases

Steffen Volz

University of Stuttgart, Institute for Photogrammetry
Geschwister-Scholl-Str. 24D, 70174 Stuttgart
steffen.volz@ifp.uni-stuttgart.de

Abstract

The fact that mobile users of location-based services (LBS) have to be able to go anywhere on earth without changing the application brings up a major requirement: there has to be one global platform that provides a trans-border and continuous access to the information necessary for this kind of information systems – and the most important information source for location-based services is geospatial data. However, the world of geospatial data is split into pieces, and the integration of those pieces is a difficult task since the data are highly heterogeneous: they have been acquired according to differing conceptual schemas (or application perspectives), they are available in different formats and scales, with different accuracies, etc. The main problem concerning spatial data integration results from the fact that the same real world objects are stored in multiple, inconsistent representations (MRep) in different spatial databases. Thus, in order to achieve a common view on the underlying spatial data within a global information platform for location-based services, mainly the inconsistencies between multiple representations have to be considered and dealt with adequately.

The problem can be depicted by means of a navigation service. Consider the task of finding the shortest car route between Stuttgart/Germany and Vienna/Austria. You will not have any problems to find a solution, if you have one single, continuous street data set comprising the whole query

area. However, you will encounter difficulties in case you have two or more separate, partly overlapping source data sets (or patches) that you have to assemble (like a mosaic) in order to form the query area. In this case your navigation application has to be able to merge (or conflate) the source data sets that contain multiple conflicting representations of one and the same real world (street) object in the overlapping areas to again create one consistent, single representation (SRep) street database on which the navigation algorithms can operate. The merge operation, though, is time consuming. This paper describes an approach to find an optimized solution for the navigation in MRep street networks by generating explicit relations between multiple representations and by exploiting them during shortest path search.

1 Introduction

The Nexus project that is currently being carried out at the University of Stuttgart [7] aims at developing a global and generic platform for different kinds of location-based services (LBS). It is based on the idea of providing an integration schema for spatial data, the so-called Augmented World Schema (AWS), into which all the available data sources of the highly distributed spatial database servers all over the world can be converted. If the data are finally formatted according to the AWS standard and made available within the Nexus platform, the problem of multiple representations (MRep) of spatial data arises since different data providers might have captured the same real world entity. As a solution, we propose to link corresponding spatial representations existing in different data sets by explicit relations, so-called MRep Relations. These MRep Relations store, amongst others, measures to express the geometric, topological and thematic similarity (or consistence) of the representations that have been assigned to each other during a matching process.

MRep Relations can on the one hand be used to support update processes within the Nexus platform and on the other hand to merge (or conflate) data sets containing multiple representations into one resulting, consistent data set that merely contains one single representation for each real world object (SRep). On such single representation data sets, conventional spatial algorithms necessary for LBS can work. Conflation, however, even of small areas, is a time-consuming process and it is illusive that a consolidated data set can be created for each client query, not to mention the efforts to produce one consistent data set for the whole world and to update it regularly. Therefore, other mechanisms have to be invented.

In this paper, we suggest that MRep Relations can also be used to avoid the conflation of corresponding objects into a single representation by developing algorithms that are able to exploit the information stored in MRep Relations during the analysis process. Since navigation is one of the most important services for location-based systems, we verify our approach by means of this application and describe a shortest path algorithm that can be used in multi-representation databases (i.e., to solve scenarios like the one sketched in the abstract). For our investigation, we use street data sets from different spatial databases, namely Geographic Data Files (GDF) [4] and the German Authoritative Topographic Cartographic Information System (ATKIS) [11] which have been transformed into the AWS format. GDF is an international standard that has especially been developed for car navigation applications and thus is used to describe and transfer street networks whereas ATKIS is more comprehensive and contains data of different topographic categories like settlement, vegetation, traffic, etc. Both data models capture street objects as linear features in approximately the same scale (1:25 000).

The remainder of this paper is organized as follows: section 2 describes related work in the field of matching and conflation of multi-representation databases. Section 3 briefly introduces the Nexus project and in section 4 the idea of MRep Relations is presented. Section 5 contains a detailed description of the navigation approach in multi-representation databases. Finally, section 6 summarizes this paper and gives an outlook on future issues.

2 Related Work

Interoperability of spatial databases is one of the main research topics in the GIS domain. Amongst others, it comprises the field of integrating multiple, potentially contradictory representations of one and the same real world object.

In order to identify corresponding representations in different databases, matching techniques are developed. For example, [2] has implemented an approach for matching street network nodes of two different GDF datasets which have been acquired by different companies (NAVTEQ and TeleAtlas). The algorithm developed here is based on the idea of describing intersections of streets, i.e. nodes of a street network, by a code. This code consists of point coordinates and the number, abbreviations and names of incident streets. For each intersection, such a code is created. By comparing the codes of the intersections within the different GDF data sets and by assigning the intersections with the most similar codes to each other, references can be derived.

A fundamental, line-based matching approach for street network data of ATKIS and GDF has been presented by [10]. In a first step, the algorithm finds all potential correspondences of topologically connected line elements in two source data sets by performing a buffer operation. The matching candidates are stored in a list. This list is ambiguous and typically contains a large amount of *n:m* matching pairs. Then, unlikely matching pairs are identified and eliminated using relational parameters like topological information and feature-based parameters like line angles. The result is a smaller but still ambiguous list with potential matching pairs. These matching pairs are evaluated with a merit function in order to compute a unique combination of matching pairs which represents the solution of the matching problem. This is a combinatorial problem which is solved with an A* algorithm.

The problem of spatial data fusion or conflation is for example being tackled by [3]. The merging process is defined here as "feature deconfliction", where all parts of a matched feature pair are unified into a single "better" feature. The conflation algorithm has to decide, which properties are preserved in the resulting instance. In their approach, the authors are also taking into account the data quality information of the corresponding instances in order to derive SRep data. Also in [5] it is dealt with the fusion of multi-source vector data. The integration process is divided into a "linking" phase where homologous objects are identified and a "best map" phase that assembles the linked objects. During best map, so-called connectivity vectors are placed at the boundaries of linked objects in order to connect the geometries of corresponding representations. Similarly to our approach, a "best path" application was developed that allows finding shortest paths between low-resolution and high-resolution street data sets. However, unlike in our work, linking does not produce any information about the degree of similarity of the linked objects that could be exploited during data analysis.

3 A Short Introduction to Nexus

The goal of the Nexus project is to provide a platform for different kinds of location-based applications. This platform is organized as a federation that provides the Augmented World Schema (AWS) – a spatial integration schema, i.e. a common conceptual schema especially designed for LBS – into which the data of autonomous spatial databases (so-called context servers, CS) have to be transformed [8]. This involves the solution of semantic issues (see [9]). Just like in the WWW, anyone who wants to participate in the Nexus system can provide an own context server that merely has to be registered at the Area Service Register (ASR) component, a spatial lookup service that is utilized to find the data necessary to answer a client query. Anytime a new context server is registered, the MRep Builder is triggered to find corresponding representations in the spatial databases already available in Nexus and in the newly added one and generates MRep Relations between them. All MRep Relations are stored in special context servers, the MRep Relation Servers that also have to be registered at the ASR. If an LBS client query has to be processed, the Federation finds out via the ASR which spatial data servers and MRep Relation Servers have to be accessed to gather the query relevant data. Those services which are capable of using MRep Relations within their analysis algorithms (MRep Services) do not need any further integration steps and can create their version of a Unified Context Model. Other services, like e.g. a map service, which need to produce a SRep data set (Non-MRep Services), exploit MRep Relations in the merging process to assemble the different context models (or patches) and create another version of a Unified Context Model. Finally, the results produced by the services can be returned to the application (see Figure 1).

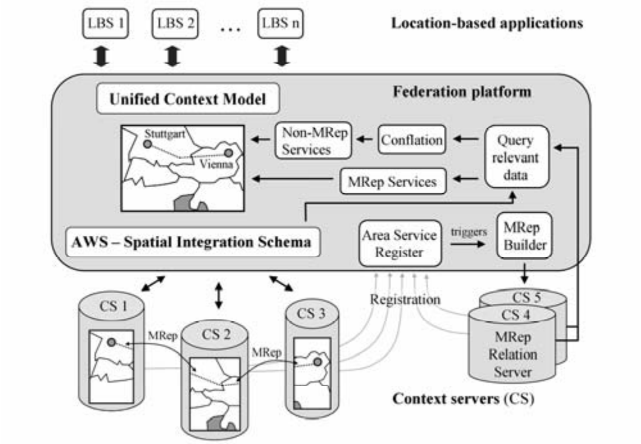


Fig. 1. Architecture of the Nexus platform

4 The Concept of MRep Relations

The first step in integrating heterogeneous databases consists of feature matching [3], i.e. representations of the same real world object within different databases have to be identified. The matches between corresponding instances can basically be stored as simple pointers within a bidirectional list, displaying that an object O_{Ak} or an object set $\{O_{A1, \ldots,} O_{Ak}\}$ of data set A can be assigned to an object O_{Bn} or an object set $\{O_{B1, \ldots,} O_{Bn}\}$ of data set B (e.g. $O_{A1} \leftrightarrow O_{B1}$, $\{O_{A3}, O_{A5}\} \leftrightarrow O_{B2}$, $\{O_{A6}, O_{A9}\} \leftrightarrow \{O_{B4}, O_{B7}\}$, etc.). However, during matching more information like geometric, topologic and attributive similarity measures for corresponding representations can be derived. This additional information is according to our approach represented in an own formal structure, the MRep Relation, which can be exploited for multiple purposes. One of these purposes is the shortest path search in multi-representation databases.

4.1 MRep Relations

MRep Relations can contain several structural elements like their position (in order to allow a spatial search of MRep Relations), identifiers of related objects, cardinality of the matches (1:1 … n:m), comparisons of data quality parameters, etc. With respect to our work, only the similarity measures that are created for corresponding representations are of relevance. Since the presented application focuses on the investigation of linear street data and potentially useful attributive information (like street names) was not available, only geometric and topological similarity indicators were generated for the linear representations to be matched. The assessment of geometric similarity was based on the comparison of angle and length differences and distance values like the average line distance and the Hausdorff distance. Adjacency relations (node degree differences) of corresponding street representations were used to detect the topological similarity. The different partial similarity measures were eventually aggregated into a total similarity value using a simple weighted sum approach: the absolute values of the individual similarity measures were first divided into 7 classes from 0 (lowest similarity) to 6 (highest similarity) to obtain so-called evaluation values. Then, each evaluation value was weighted with a factor. The weight factors were specified on the basis of the operator’s expertise regarding the influence of the different partial similarity values on the total similarity. The total similarity value was normalized onto an interval ranging from 0 to 100, with 100 representing the highest similarity.

MRep Relations are stored and exchanged using an XML-based format called **M**ulti**R**epresentational **R**elation **L**anguage (MRRL) which was specified within this work. The following extract of MRRL shows the basic structure of an MRep Relation (see Figure 2).


```
<mreprelation>
 <mrep_id /> <position />
 <attributes>
  <general_atts>
   <source_ids>atkis_awml.349;</source_ids>
   <target_ids>gdf_awml.826;gdf_awml.827;</target_ids>
   <cardinality>1:2</cardinality>
   <total_similarity>90.56</total_similarity>
  </general_atts>
  <semantic_atts />
  <geometric_atts>
   <length_difference>13.84</length_difference>
   <avg_line_distance>8.83</avg_line_distance>
   <hausdorff_distance>11.45</hausdorff_distance>
        ...
  </geometric_atts>
  <topologic_atts>
   <startnode_deg_diff>0</startnode_deg_diff>
   ...
  </topologic_atts>
        ...
 </attributes>
</mreprelation>
```


Fig. 2. Part of an MRRL instance file displaying some structural elements of a MRep Relation

4.2 Generating MRep Relations

For a test area in the inner city of Stuttgart/Germany, two different street databases originally stemming from ATKIS and GDF have been transformed into the exchange format of Nexus. Then, after reducing the global geometric deviation of the data sets by a rubber-sheeting transformation, MRep Relations have been created for the multiple representations in the overlapping areas of the two data sets. For the acquisition of MRep Relations, a semi-automatic software, the so-called Relation Builder Toolbox, has been developed (see Figure 3). It has been implemented within the framework of the publicly available Java-based GIS environment JUMP [6]. Using this tool, a human operator can manually select the instances of corresponding representations by analyzing geometric and topological criteria. For the matching, the operator is provided with a set of rules that are necessary in order to achieve replicable results amongst different operators (see [9] for further details). If the corresponding representations have been selected, MRep Relations (including the similarity values) are created automatically by the software. Currently, we are working on automated matching algorithms to minimize the intervention of human operators.

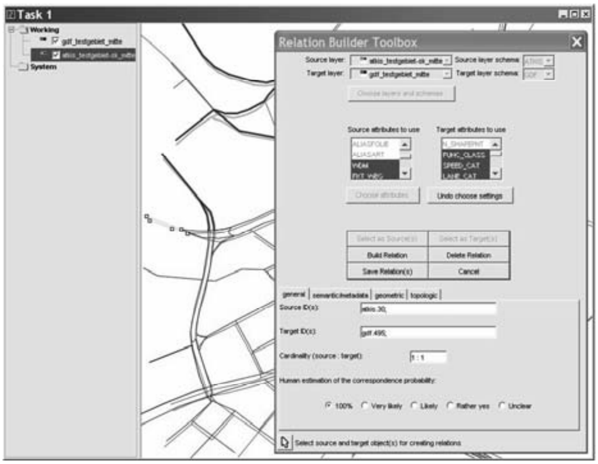


Fig. 3. The MRep Relation Builder Toolbox allows a manual matching of corresponding street representations

5 An Approach for Shortest Path Search in MRep Databases

In the following sections, the steps to achieve a shortest path search in MRep databases are presented and the developed sample application is briefly outlined.

5.1 Generating an integrated navigation graph

The principal goal of the navigation approach in multi-representation databases is to find the shortest path from a starting node in street network (or graph) I to a target node in street network (or graph) II. Switching from graph I to graph II is allowed arbitrarily often. The situation is illustrated in Figure 4 where two small, undirected graphs are displayed. From both graphs three edges and four nodes are available as multiple representations (MRep Edges and MRep Nodes). MRep Relations were only created for the edges, but not for the nodes, i.e. similarity measures are only available for the MRep Edges.

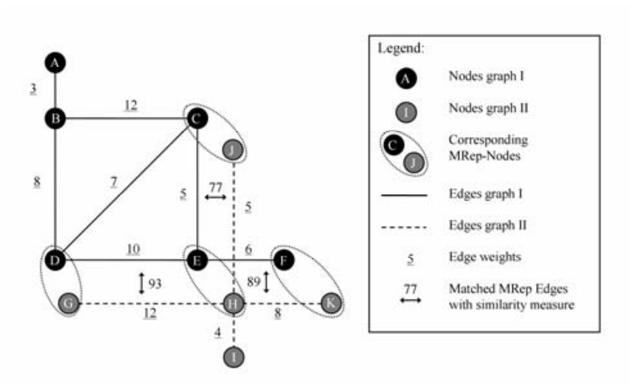


Fig. 4. A multi-representation graph

In order to create *one* consistent navigation graph from the two underlying source graphs, two options are possible: First, MRep Nodes and MRep Edges could be topologically (not geometrically) aggregated to derive the final navigation graph. However, using this option the information stored in the similarity measures would get lost. For this reason it was decided to realize the second option by introducing so-called transition edges from one graph to the other between MRep Nodes (see Figure 5). Thus, it is possible to determine the uncertainty that arises when a switch-over from one graph to the other is performed. The uncertainty results from the fact that two MRep Nodes might in reality not be identical which is expressed by the similarity measure.

5.2 Determining Possible Paths

Algorithms for shortest path search can be based on adjacency matrices like e.g. the Dikstra or Floyd algorithm. However, an adjacency matrix can also be transformed into a tree structure where the start node is represented as the root. On tree structures, several search algorithms can operate (see

[1] for an overview). We decided to apply a depth-first search to find all possible ways between a source and a target node.

5.3 Calculating the Costs

When all possible paths within the integrated navigation graph have been determined by a depth-first search, the total costs of the traversed edges (C_W) can be calculated. It can be defined as the sum of the costs of all real edges (C_{RE}) plus the sum of the costs of all transition edges (C_{TE}).

$$C_W = \sum C_{REi} + \sum C_{TEk}$$

Remember that multiple switch-overs from one graph to the other are allowed, that's why there can be more than one transition edge. For the calculation of the costs for the traversal of transition edges it has to be considered that there are only similarity measures available between corresponding multiple representation edges (MRep Edges), but not for their start and end nodes (the MRep Nodes). However, to determine the cost of a transition edge, the similarity measures for the MRep Nodes have to be known. Therefore, the similarity values of MRep Edges have to be converted to similarity values of the MRep Nodes and then the costs for transition edges can be derived.

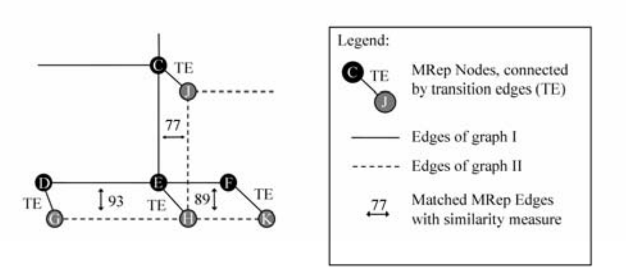


Fig. 5. The costs of transition edges have to be calculated regarding the similarity values of the MRep Nodes

In those cases where MRep Nodes only have one incident MRep Edge (e.g. MRep Nodes C and J in Figure 5), we can simply use the similarity measure of the incident MRep Edges (CE and JH = 77) as the similarity measure of the MRep Nodes. Whenever MRep Nodes do have more than one incident MRep Edge (like E and H), we chose the arithmetic mean value of the similarity values of all incident MRep Edges (for E,H = (77 + 89 + 93) /3 = 86,33) as an acceptable approximation for the similarity of the respective MRep Nodes.

The costs of each transition edge (C_{TE}) can finally be calculated by a simple cost function

$$C_{TE} = (Sim_{max} - Sim_{TE}) * \omega$$

where Sim_{max} is the maximum similarity value between MRep Nodes (100) and Sim_{TE} is the similarity value of the investigated MRep Nodes (e.g. 86,33 for E and H), i.e. the higher the similarity of MRep Nodes (Sim_{TE}), the lower the cost for traversing the transition edge. The influence of the transitions during the shortest path calculation can be weighted by a factor ω (with $\omega \in [0..n]$) and so the uncertainty which results from switching graphs can be expressed to a degree necessary for the application. The value of ω has to be chosen based on empirical observations and depends on the data sets involved. One strategy to determine a reliable result of the shortest path search could also be to neglect all paths containing transition edges with a cost above a certain threshold (e.g. above 15 for $\omega = 1$).

5.4 Realization of the Approach

The approach has been implemented as a sample application called MRep Network Analysis Tool within the JUMP environment. The user can select a start and a target node in either of the two source street networks which originally stem from GDF and ATKIS and have been translated in the AWS format. Depending on the ω value of the cost function, there can be different results for the shortest path (see Figure 6, ways A and B). Concerning way A, ω is large and so the cost function for the transition edges yields large values since the application requires a high reliability of the resulting way, i.e. the application demands that the transition has a strong influence on the length of the shortest path. Therefore, the actual transition rather takes place at those MRep Nodes that show the highest similarity (at transition A), i.e. at the place where the transition costs are minimal. As it is depicted in Figure 6, this constraint leads to a considerable detour. In the case where the transition takes place at B, the costs for the transition edges are comparably low (since ω is low), i.e. the transition has only a small influence on the length of the shortest path since the application allows a certain degree of uncertainty of the resulting path.

The reason for the uncertainty in the described example results from the fact that the MRep Edge with GDF origin might be a dead-end street at transition B (although all streets directly at the boundary appear as dead-ends), whereas the corresponding ATKIS representation is connected to other roads at transition B.

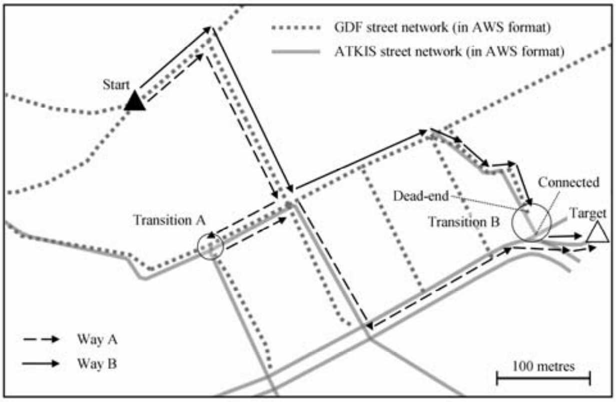


Fig. 6. Clipping of a boundary zone of different street networks. Between the MRep Edges, MRep Relations have been defined.

6 Summary and Outlook

In this paper we showed that explicitly linking corresponding objects in multi-representation databases through MRep Relations is an alternative to the conflation of multiple representations into a single representation (SRep) data set since GIS algorithms can be adapted to exploit MRep Relations during data analysis. We have verified our approach by means of a shortest path search in multi-representation street databases: First, we created MRep Relations between corresponding street objects (MRep Edges) within the originally separate GDF and ATKIS networks. Secondly, in order to connect the graph structures derived from the street networks, we introduced transition edges between MRep Nodes. The costs for traversing these transition edges depend on the similarity values stored in MRep Relations and can be further adjusted according to the reliability requirements of the applications. Thirdly, we applied a depth-first search algorithm that allows finding the shortest path between any two nodes in the unified graph. The approach was finally implemented in a sample application.

In the future, some of the processes that are part of our approach have to be extended and enhanced. Currently we are working on algorithms to automate the generation of MRep Relations as far as possible. Furthermore, issues of optimizing the tree search are being dealt with. Eventually, the creation of MRep Relations has to be adapted to other object and geometry types (e.g. buildings represented by aerial geometries) and it has to be investigated how other GIS algorithms like buffering, intersection, etc. can be extended to be able to exploit MRep Relations in the analysis process.

Acknowledgements

The research presented here is supported within a Centre of Excellence called “Spatial World Models for Mobile Context-Aware Applications” under grant SFB 627 by the German Research Council. The test data have kindly been provided by the NAVTEQ company and the state survey office of the federal state of Baden-Württemberg.

References

[1] Aho, A., Hopcroft, J. E., Ullman, J. D. (1987): Data Structures and Algorithms, Addison-Wesley Series in Computer Science and Information Processing, 427 p.
[2] Bofinger, J.M. (2001): Analyse und Implementierung eines Verfahrens zur Referenzierung geographischer Objekte. Diploma Thesis at the Institute for Photogrammetry, University of Stuttgart, unpublished, (2001), 76 p.
[3] Cobb, M., Chung, M., Miller, V., Foley, H., Petry, F., Shaw, K. (1998): A rule-based approach for the conflation of attributed vector data, GeoInformatica 2(1), pp. 7-35.
[4] Comité Européen de Normalisation (CEN) (1995): Geographic Data Files Version 3.0, TC 287, GDF for Road Traffic and Transport Telematics.
[5] Edwards, D., Simpson, J. (2002): Integration and Access of Multi-Source Vector Data. In: Proc. of the Joint International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, on CD-ROM.
[6] Java Unified Mapping Platform JUMP, 2005: http://www.jump-project.org/
[7] Nexus project homepage: http://www.nexus.uni-stuttgart.de
[8] Nicklas, D., Mitschang, B., “The Nexus Augmented World Model: An Extensible Approach for Mobile, Spatially-Aware Applications”, Proc. of the 7th Int. Conf. on Object-Oriented Information Systems, Calgary, 2001.
[9] Volz, S. (2005): Data-Driven Matching of Geospatial Schemas. In: D. Mark and A. Cohn (eds.): Proceedings of the Conference on Spatial Information Theory (COSIT '05), Ellicottville, NY, Lecture Notes in Computer Science 3693, Springer, September 2005, pp. 115-132.

[10]Walter, V., Fritsch, D. (1999): Matching spatial data sets: a statistical approach. Int. J. Geographical Information Science 13(5), pp. 445–473.

[11]Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (1988): Authoritative Topographic Cartographic Information System (ATKIS), http://www.atkis.de

Region-Based Representation for Assistance with Spatio-Temporal Planning in Unfamiliar Environments

Inessa Seifert, Thomas Barkowsky, Christian Freksa

Universität Bremen
{seifert, barkowsky, freksa}@sfbtr8.uni-bremen.de

Abstract

In this contribution we present a cognitively motivated approach to an interactive assistance system for spatio-temporal planning tasks. Since mental spatial problem solving is known to be based on hierarchical representations, we argue for region-based representation structures that allow for structuring a complex spatio-temporal problem such that exploring and communicating alternative solutions to a given problem are easily enabled. In this way, spatio-temporal constraints can interactively be dealt with to find appropriate solutions for a given planning problem.

1 Motivation

Ubiquitous assistance with spatial problem solving tasks like wayfinding from specific locations to specific other locations gained much attention in spatial cognition research (e.g., Richter & Klippel, 2005, Denis, 1997; Tversky & Lee, 1999). However, early stage route planning, i.e., pre-visiting of locations using representations of unfamiliar large-scale environments like maps has been considered in the literature only seldomly (e.g., Dirlich et al., 1986).

Planning a round trip or a city tour in a foreign country are good examples for such type of problems. The problem solving task aims at selecting activities and putting them into a feasible order under consideration of temporal and spatial constraints. Temporal constraints include, for example, predefined durations of the whole trip and durations of individual activities. Spatial constraints include partially specified location assignments and distances between locations where the activities take place. Especially in an early route planning stage, assignments of locations are specified at different levels of granularity. Location assignments can be specific places of interest (e.g., a particular museum), or regions (e.g., parts of a country, national parks, or cities). In other cases, since the environment is unfamiliar, there are no particular location assignments, but rather specific requirements on properties of the corresponding locations resulting from the type of the planned activities (e.g., swimming, hiking, or sightseeing). To solve the illustrated partially underspecified spatio-temporal planning problem means to find all possible location assignments for activities with unspecified location assignments and to obtain feasible spatio-temporal configurations which fulfill the given temporal scope and spatial constraints.

The illustrated problem solving task can be formalized as a spatio-temporal constraint satisfaction problem (cf. Dechter, 1992). However, taking into consideration all possible assignments of locations left unspecified the demonstrated task 1) has a high computational complexity, and 2) results in a large solution space. To reduce computational complexity, the search process should be guided by a human user. In doing so, significant parts of the problem space can be omitted. Yet, allowing users to participate in the search process limitations of human cognitive capacity have to be taken into consideration. Observation and comparison of a large number of alternative solutions is a cognitively demanding task (cf. Knauff et. al., 2002). Consequently, the problem space has to be adequately represented to a user in order to reduce the mental load occurring during observation and comparison of alternative solutions.

Referring to the definition by Simon (1981), "*… solving a problem simply means representing it so as to make the solution transparent*", we have to understand how humans conceive spatial and temporal information and how the intermediate problem solving states can be adequately communicated to a user. The paper describes an approach for structuring and representing a spatio-temporal problem domain which facilitates the generation of alternative solutions and provides a communication basis for collaboration of a person and an artificial assistance system.

2 Collaborative Assistance with Spatio-Temporal Planning

To obtain feasible solutions for the given spatio-temporal configuration problem, the constraints can be formalized as a *constraint network* (cf. Dechter, 1992). The constraint network consists of a finite set of variables, e.g., a set of activities *Conf* = $\{a_0, a_1, \ldots, a_i\}$, where $0 \leq i < \mathbf{n}$ and **n** is the number of activities, $\mathbf{n} \in N$. Each activity has the following attributes: *type, duration*, and *location*. Regarding the assignment of the attributes, duration is mandatory, whereas assignments of location and activity type are optional. Let the temporal interval **m** represent the duration of the whole journey. A duration of an activity a_i can be obtained by the duration function $dur\_cost(a_i)=d_i$, where $d_i \neq 0$, $d_i \in R$. An activity type is an element from $Type = \{t_0, t_1, ..t_k\}$, $k \in N$, where k is a number of activity types. A location can be a geographical position, i.e., a point with the corresponding geographical coordinates (x, y) on a 2-dimensional geographical map where one or more activities can be accomplished, or a region. Each location is associated to one or more activity types. Locations are connected with each other by paths, which have distance costs.

The constraints which may hold between the corresponding activities are expressed by the thirteen qualitative temporal relations described in (Allen, 1983). Distances between locations of activities, i.e., a function $dist\_cost(a_i, a_{i+1})$ are represented as temporal costs, i.e., time required to get from one location to another. Each feasible spatio-temporal configuration *Conf* = $(a_0, a_2, \ldots, a_i)$, where $0 \leq i < \mathbf{n}$ and **n** is the number of activities, should fulfill the following constraints: (1) The temporal relations which holds between subsequent activities a_i, a_{i+1}, where $0 \leq i <$ **n**, should be “before” (*b*) or “meets” (*m*). (2) The sum of durations and the temporal distance costs between the corresponding locations should not exceed the given temporal scope of a journey **m**, i.e., $\Sigma\, di + \Sigma\, dist\_cost(a_i, a_{i+1}) \leq \mathbf{m}$. (3) Underspecified spatial constraints should result in different location assignments. To obtain all feasible spatio-temporal configurations we need to search through all considerable location assignments.

2.1 Need for assistance

Since the given environment is unfamiliar and we have to deal with a large number of activities, it is difficult to solve the illustrated problem mentally. Depending on the number of unspecified location assignments the demonstrated problem is weakly constrained. Underspecified activities contribute to large solution spaces. Consequently, the provided alternative spatio-temporal configurations require mental post-processing, i.e., they have to be observed by a user in order to find a solution which corresponds to the user's specific interests and preferences.

2.2 Partially unformalized constraint systems

The demonstrated problem can be considered as a partially unformalized constraint problem (Schlieder & Hagen, 2000). The example problem solving task described in (Schlieder & Hagen, 2000) is a geometric layout constraint satisfaction problem. To provide assistance with such type of problems, they separated the constraints into formalized *hard* geometric *constraints*, e.g., ordering and alignment of elements in a spatial configuration, and s*oft constraints* like personal aesthetic preferences, which cannot be formalized. The assistance system provides users with a set of geometric configurations which fulfill the specified *hard* alignment *constraints*. Subsequently, users have to examine the generated solutions to obtain a "good" solution, which fulfills their personal requirements on solution quality regarding the personal *soft constraints*. To facilitate the mental coprocessing during the interactive search process, the introduced assistance system adapts to human reasoning strategies regarding alternative spatial configurations, i.e., reasoning with mental models (Johnson-Laird, 1983), and particularly reasoning with preferred mental models (Knauff et al., 1995).

In this vein, the demonstrated spatio-temporal planning problem can be treated likewise as a partially unformalized constraint problem, consisting of predefined spatial and temporal constraints and a user's personal interests and wishes to spend his/her time in a particular location.

However, in contrast to Schlieder and Hagen's geometric layout-problems, depending on the number of activities, we may have to deal with a greater number of objects, which have to be observed in each spatio-temporal configuration. Calculations of all possible configurations including various permutations in order of activities require much computational resources of the assisting devices. Consequently, we need a compact representation of the problem domain which on the one hand facilitates the generation of alternative solutions and on the other hand adapts to mental reasoning strategies about spatio-temporal relations.

2.3 Dealing with the complexity of geographic information

To master the computational complexity of the given spatio-temporal configuration problem we have to single out particular aspects of the problem domain, i.e., objects and specific properties and relations between them which are relevant for the problem solving task and make it possible to solve the problem efficiently (cf. Freksa & Barkowsky, 1996). Humans deal with spatial problems like route planning surprisingly well – especially in familiar environments. Therefore, organizational and structural principles of mental representations gained much attention in disciplines like cognitive psychology, artificial intelligence, and particularly in robotics. In this vein, the derived organization principles for structuring spatial information can be applied for the knowledge representations in artificial spatial systems.

The cognitively motivated principles for structuring representations of large-scale environments allow on the one hand for managing the computational complexity and on the other hand for establishing an adequate dialog between a human and an assistance system. In the following, we will describe how spatial knowledge about large-scale environments is structured in human memory and how humans operate on their mental representation when solving such types of spatial problems.

3 Hierarchies in Human Mind

Mental spatial knowledge representations are often described as cognitive maps (cf. Hirtle & Heidorn, 1993), cognitive atlases (cf. Hirtle, 1998), or cognitive collages (Tversky, 1993). Evidences for hierarchical representations of spatial information in human mind can be found in, e.g., (Hirtle & Jonides, 1985), (Tversky, 1993), and (Stevens & Coupe, 1978). Spatial knowledge stored in human long-term memory is often described as distorted, consisting of various loosely coupled knowledge fragments which are incomplete, inconsistent, and even contradictory (cf. Tversky, 1993). Nevertheless, humans perform surprisingly well when solving spatial problems like route planning and wayfinding in familiar environments. On that account much scientific research efforts were put into investigations on how humans learn, structure, and reason about familiar and partially unfamiliar spatial environments in order to obtain the main aspects of spatial information that are utilized for spatial problem solving tasks.

3.1 Structural aspects of spatial environments

The pioneering work of Lynch (1960) describes how the citizens of three different American cities structure the information about their urban environment, i.e., the mental images of their cities. Lynch defines the main structural elements that result from his various exploration studies as *landmarks*, *nodes*, *paths*, *edges*, and *districts*. *Landmarks* are salient features of the urban environment like buildings and places. They serve the observer as reference points. *Paths* are streets or lanes. *Nodes* are path intersections, bridges, or changes of transportation modes and form important points. *Edges* are boundaries of areas, which form a physical barrier (e.g., rivers, highways from a pedestrian point of view). *Districts* are areas in a city that share a common property (e.g., shopping areas, residential areas). Lynch describes that individuals tend to use global landmarks in new environments for the purpose of orientation. Gradually, the individuals add further information to their knowledge until a cognitive map is constructed.

Consequently, mental representations consist primarily of topological knowledge, i.e., how the landmarks are interrelated. Depending on the medium used for acquiring knowledge about the environment such topological representations can be gradually mapped to survey knowledge, allowing for adding new spatial relations to cognitive maps.

3.2 Human spatial problem solving strategies

When solving spatial problems, humans operate on hierarchical knowledge representations stored in long-term memory. Stevens and Coupe (1978) demonstrated the hierarchical organization of mental geographic knowledge using a spatial problem, i.e., asking students of a Californian university about relative positions of two American cities, Reno and San Diego. The results of the experiment have shown that the subjects derived the spatial relations between two cities from the super-ordinate spatial relations between the states of California and Nevada when solving this problem. To obtain the missing knowledge about the relations between the two cities, subjects utilized the containment relations of the cities to the corresponding states, which resulted in most of the cases in a false answer, i.e., Reno being further east than San Diego, albeit it is further west (see figure 1).

Fig. 1. Map of California and Nevada exhibiting the relative positions of the cities of San Diego and Reno

Hierarchies in human mind are used for categorization of information and particularly for structuring large-scale environments into regions. Regions "*help people organize their understandings of the world in an efficient manner, they also help various activities in space occur more efficiently*" (Montello, 2003).

The principle of *regionalization* of large-scale environments provides a basis for series of navigation experiments conducted in a virtual reality lab (Wiener, 2004). The results of the experiments have shown that regionalization facilitates spatial problem solving tasks like navigation in unfamiliar large-scale environments. Mental representations formed in regionalized environments contain super-ordinate connectivity relations, i.e., *region connectivity*, which allow for performing navigation tasks more efficiently.

4 Hierarchies in Artificial Spatial Systems

"*In order to effectively manage complex environments the spatial system must be capable of discarding useless information and breaking large environments into collections of smaller environments*" (Chown, 2000). Therefore, a significant body of research is concerned with development of spatial systems which operate on hierarchically organized spatial knowledge representations (e.g., Chown, 2000; Kuipers, 2000; Leiser & Zilberschatz, 1989).

Mental representations contain primarily topological information which describes how different locations are interconnected. This outstanding property of mental representations was successfully applied in various artificial spatial systems, making their orientation and route planning performance more robust. For example, the topological level of Kuipers' Spatial Semantic Hierarchy (2000) consists of the features of external environment, defined by places, paths, regions and their connectivity, order, and containment relations. Also the spatial system PLAN by Chown (2000) operates on topological representations which consist of a network of topologically interrelated landmarks and a network of gateways. The gateways mark the transitions between different regions of space.

Spatial systems have to be capable of structuring the spatial knowledge into regions and be able to distinguish between the paths which connect places from those connecting regions. Although hierarchical route planning algorithms provide sub-optimal solutions, they require less computational power and are more efficient.

5 Regionalize and Conquer

In the following, we are going to introduce a representation of the spatio-temporal problem domain which utilizes the demonstrated properties and aspects of hierarchical spatial knowledge representations in artificial systems as well as in human mind.

To establish an adequate dialog between a human and an assistance system we have to structure the problem domain into regions. Operations on hierarchical region-based structures resemble human hierarchical spatial reasoning strategies (Hirtle & Jonides, 1985; Stevens & Coupe, 1978). Furthermore, "*regionalization has its definite analytic and communicative utility. It simplifies complexity and avoids unnecessary precision, both in thought and speech*" (Montello, 2003). Consequently, the proposed spatial representation consists of *locations*, *regions*, *nodes,* and *paths*.

Locations are points of interest, where one or more types of activity can take place. A location is an abstract one-dimensional point. To communicate a location to a user, each location has a name. The name has to be unique within an activity region that contains the location. Locations can also serve as an abstraction of a region.

An *activity region* contains one or more spatially proximate locations and a connectivity node. The locations within an activity region share specific properties according to the type activities, which are supposed to be pursued in that region.

A *connectivity node* is a location, which connects a set of locations within an activity region (see figure 2). Such node has a connectivity cost, which results from a connectivity measure, e.g., an average distance cost to the corresponding locations.

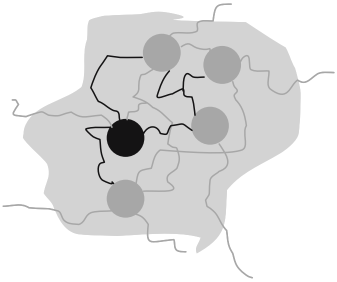


Fig. 2. A node stands for connectivity costs of the locations within a region

Super-ordinate regions structure an unfamiliar environment and contain activity regions. The structuring criteria encompass properties of the environment which are relevant for the current spatio-temporal planning task. For example, when planning an individual journey administrative regions, e.g., federal states together with topography can be considered as relevant. The containment relation is represented by a *part-of* relation between super-ordinate und activity regions.

The proposed representation structure consists of a hierarchy of locations, activity regions and super-ordinate regions, related to each other by *part-of* relations. Such partonomies allow for representing and reasoning about qualitative topological relations between regions with rough boundaries (Bittner & Stell, 2002). The connectivity nodes assigned to activity regions are linked with each other via paths. Paths consist from nodes and edges and carry the distance costs.

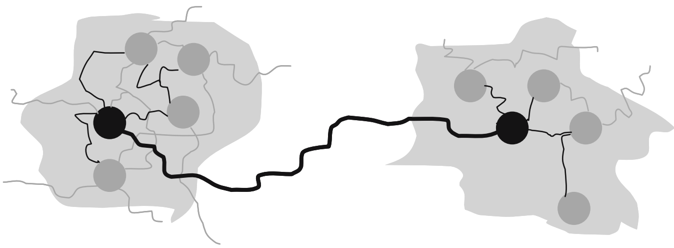


Fig. 3. Nodes connect regions

Figure 3 illustrates different locations within a region. One of the locations represents a connectivity node that binds other locations. Such node can be one of the locations of interest, or another alternative location that binds the locations within a region with a better connectivity cost.

The proposed hierarchical region-based representation is built in a bottom-up manner. The following section describes how the proposed region-based hierarchy allows for solving spatio-temporal planning problems efficiently in an interactive way.

5.1 Region-based assistance with spatio-temporal planning

It is assumed that the required information about the locations of interests and paths between them can be provided to the assistance system by a geographic information system (GIS). The assistance system first generates the corresponding activity regions that contain several locations of interest of the required activity type.

Figure 4 shows an abstract depiction of a hierarchically organized region-based representation. The locations of interest are bound into activity regions associated with the activity types: “hiking”, “swimming” and “sightseeing”. The locations inside the activity regions are bound to the corresponding connectivity nodes: *node 1a, node 2a, node 3a, node 4a, node 5a,* and *node 1b, node 2b, node 3b, node 4b*, respectively. Each of these connectivity nodes represents a connectivity cost for the corresponding activity type within an activity region. *Region A* and *Region B* contain the sets of regions {*1a, …,5a*}, {*1b,…,4b*}.

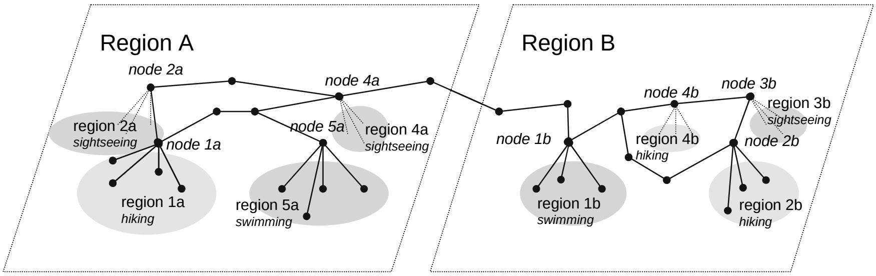


Fig. 4. Region-based representation

In order to obtain all feasible spatio-temporal configurations all permutations in the order of the activities in the given spatio-temporal configuration *Conf* = (a_0, a_2, …, a_i) and all possible assignments of the corresponding regions, e.g., *1a, 2a, 3a* etc. have to be taken into consideration. However, a significant part of the problem space can be pruned when using partial orders of the super-ordinate regions *{Region A, Region B}* or *{Region A, Region B}.*

A user guides the search process by definition of additional constraints, e.g., selection of partial orders of super-ordinate regions, or partial orders in the given set of activities. The proposed representation allows on the one hand for specification, on the other hand for relaxation of constraints at different levels of the spatial region-based hierarchy. Furthermore, assistance with pre-visiting of unfamiliar environments can be provided at different levels of granularity with no requirement on precise assignments of particular locations, which can be refined at a later point of a journey. In this vein, the assistance system accompanies the problem solving process, resembling human hierarchical reasoning strategies about spatial relations. It provides a possibility to define partial orders on a coarse level and to step into the detailed level to find feasible partially refined spatio-temporal configurations keeping the global solution space consistent.

6 Outlook and Future Work

The introduced hierarchically organized region-based representation allows for assignment of spatial constraints and definition of partial orders of regions and activities at different levels of granularity. The proposed assistance system operates on hierarchically organized locations, activity regions and super-ordinate regions. For the purpose of the simplicity of the illustrated example, an adequate mapping procedure of activity regions to commonly used administrative regions has not yet been considered. Since activity regions may contain locations of different activity types, they may overlap with more than one administrative region. In order to deal with such cases structuring criteria based on qualitative spatial relations between activity regions and administrative regions have to be considered. However, this is one of the topics of future work.

Another important topic is the establishing of an adequate interaction model between an assistance system and a user. Skillful navigation operations for observation of the solution space have to be defined. Such operations serve for representation and manipulation of the spatio-temporal configurations, like, e.g., the demonstrated specification of partial orders at different levels of the region-based hierarchy. For this purpose human problem solving strategies (cf. Newell, & Simon, 1972) and reasoning with mental models (cf., Johnson-Laird, 1983), particularly preferred mental models (cf. Knauff et. al., 1995) can be utilized. These topics are also a matter of further research.

Acknowledgements

We gratefully acknowledge financial support by the German Research Foundation (DFG) through the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition (project R1-[ImageSpace]). We also thank Holger Schultheis for valuable comments and numerous discussions.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM, 26* (11), 832-843.

Bittner, T, & Stell, J.G. (2002). 'Vagueness and Rough Location'. Geoinformatica, Vol. 6, 99-121.

Chown, E., (2000). Gateways: An approach to parsing spatial domains. In ICML 2000. Workshop on Machine Learning of Spatial Knowledge.

Dechter, R., (1992) "*Constraint Networks (Survey).*" In Encyclopedia of Artificial Intelligence, 2nd edition, 1992, John Wiley & Sons, Inc., pp. 276-285.

Dirlich G, Freksa C, Schwatlo U, Wimmer K, 1986 Hg, *Kognitive Aspekte der Mensch-Computer-Interaktion*, 22-36, Springer, Berlin 1986.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. Cahiers de Psychologie Cognitive,16,409-458.

Freksa, C., & Barkowsky, T. (1996). On the relation between spatial concepts and geographic objects. In P. Burrough & A. Frank (Eds.), Geographic objects with indeterminate boundaries (pp. 109-121). London: Taylor & Francis.

Hirtle, S. C. (1998). The cognitive atlas: using GIS as a metaphor for memory. In M. Egenhofer & R. Golledge (Eds.), Spatial and temporal reasoning in geographic information systems (pp. 267-276). Oxford University Press.

Hirtle, S. C., & Heidorn, P. B. (1993). The structure of cognitive maps: Representations and processes. In T. Gärling & R. G. Golledge (Eds.), *Behavior and environment: Psychological and geographical approaches* (pp. 170-192). Amsterdam: North-Holland.

Hirtle, S. C., & Jonides J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition, 13* (3), 208-217.

Johnson-Laird, P. N. (1983). Mental models. Cambridge, Harvard Univ. Press.

Knauff, M., Schlieder, C., & Freksa, C. (2002). Spatial Cognition: From Rat-Research to Multifunctional Spatial Assistance Systems. Künstliche Intelligenz, Heft 4/02, arendtap Verlag, Bremen.

Knauff, M., Rauh, R., & Schlieder, C. (1995). Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society (pp. 200-205). Mahwah, NJ: Lawrence Erlbaum.

Kuipers, B. (2000). The spatial semantic hierarchy. *Artificial Intelligence, 119,* 191-233.

Leiser, D., & Zilbershatz, A. (1989). The Traveller: A computational model of spatial network learning. *Environment and Behavior*, *21*, 435–463.

Lynch, K. (1960). *The image of the city*. Cambridge, MA: MIT Press.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), Foundations of Geographic Information Science (pp. 173-189). London: Taylor & Francis.

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice Hall.

Richter,K., Klippel (2005), A.. A Model for Context-Specific Route Directions. In Christian Freksa, Markus Knauff, Bernd Krieg-Brückner, Bernhard Nebel, Thomas Barkowsky (Eds.), Spatial Cognition IV. Reasoning, Action, Interaction: International Conference Spatial Cognition 2004, pp. 58–78. Springer, Berlin.

Schlieder, C., & Hagen, C. (2000). Interactive layout generation with a diagrammatic constraint language. In C. Freksa, C. Habel, & K. F. Wender (Eds.), Spatial cognition II - Integrating abstract theories, empirical studies, formal methods, and practical applications (pp. 198-211). Berlin: Springer.

Stevens, A., & Coupe. P. (1978). Distortions in judged spatial relations. *Cognitive* Psychology, 10, 422-437.

Simon, H.A. (1981). The Sciences of Artificial. (2^{nd} ed.), MIT Press, Cambridge.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. Frank & I. Campari (Eds.), *Spatial information theory* (pp. 14-24). Berlin: Springer.

Tversky B. and Lee P. U. (1999). Pictorial and Verbal Tools for Conveying Routes. In C. Freksa, D.M. Mark (Eds.): COSIT'99, pp. 51-64, Springer.

Wiener, J.M., (2004) (PhD Thesis, University of Tübingen, Germany): Places and Regions in Perception, Route Planning, and Spatial Memory. (2004)

A Logic-Based Foundation for Spatial Relationships in Mobile GIS Environment

Mohammad Reza Malek[1, 2], Andrew U. Frank[3], Mahmoud Reza Delavar [2]

[1] Dept. of Geodesy and Geomatics Eng., K.N. Toosi University of Technology, Tehran, Iran, malek@mashhad.ncc.org.ir
[2] Geomatics Eng. and Disaster Prevention Center of Excellence, Dept. of Surveying and Geomatics Eng., Univ. College of Eng., University of Tehran, Tehran, Iran, mdelavar@ut.ac.ir
[3] Institute for Geoinformation and Cartography, Technical University Vienna, Austria, frank@geoinfo.tuwien.ac.at

Abstract

The mobile computing is a new revolutionary style of technology that enables us to access information anywhere and anytime. Mobile GIS as an integrating system of mobile computing and some GIS capabilities has fostered a great interest in the GIS field. Although mobile computing has expanded in the past decade, there are still some important constraints the complicate work with a mobile information system. The limited resources in the mobile computing would restrict some features that are available in the traditional computing technology. This chapter attempts to provide a paradigm to treat moving objects in a mobile GIS environment. An idea based on space and time partitioning is suggested. A logic-based framework for representing and reasoning about qualitative spatial relations over moving objects in space and time is proposed. We provide convincing evidence of this theory, by demonstrating how it can provide a framework model of topological relations in space and time. The expressivity power of the proposed framework is shown with some new topological relationships between moving objects and describing the

coaching problem in a mobile environment. The latter finds its application in RoboCup championship and battlefield, as well.

1 Introduction

Mobile computing is a new revolutionary style of technology emerging of the advances in the development of portable hardware and wireless communications [23, 25]. It enables us to access information anywhere and anytime. Advances in location-based engines and on-board positioning sensors lead to mobile geospatial information systems (GIS) [27, 31]. Mobile GIS as an integrating system of mobile computing and some GIS capabilities has fostered a great interest in the GIS field [21]. It becomes a new branch of GIS and brings GIS into a new stage of development.

Although mobile computing has been increasingly grown in the past decade, there still exist some important constraints which complicate the design of mobile information systems. The limited resources in the mobile computing would restrict some features available on the traditional computing. The resources include computational resources (e.g., processor speed and memory), user interfaces (e.g., display and pointing device), bandwidth of mobile connectivity, and energy source [4, 16, 18, 19, 34, 36, 45]. In addition, it is assumed that in a mobile GIS environment, sensors on the user side could not access all relevant information about other users (not complete data) and they are mainly concerned with the user and its neighbors (not global data).

Among of the most important characteristics of qualitative properties of spatial data and perhaps the most fundamental aspect of space are topology and topological relationships. Topological relations between spatial objects are such relationships that are invariant with respect to specific transformations due to homeomorphism. The study of topological relationships is firmly evolving as an important area of research in the mobile GIS [34, 36]. Though in some respects, it closely resembles topological relationships in the traditional desktop GIS, the unique properties of the mobile GIS environment demand its careful and separate study. Hence, we pay attention to this topic as well.

In this chapter, a logical framework is presented in order to provide a paradigm that treats with moving objects in a mobile GIS environment. In this framework the concept of spatial influenceability from relativistic physics, is combined with the partition and conquer idea from computer science. It means dividing the space and time into small parts; say space-time cell; and using influenceability concept presented in this article, provides a theoretical framework of mobile objects in space-time. In our view, influenceability which stands for spatial causal relations, i.e., objects must come in contact with one another; is primary an order relation.

We provide convincing evidence of this theory, by demonstrating how it can provide a framework model of topological relations in space as well as in time. The expressivity power of the proposed framework is shown with some new topological relationships between moving objects and describing the coaching problem in a mobile environment. The latter finds its application in RoboCup championship and battlefield, as well.

The remainder of the paper is structured as follows. Section 2 reviews related works. Section 3 shall present the fundamental concepts. Section 3 introduces our suggested model. In section 4 we discuss some examples of spatio-temporal relationships between two moving agents and its application in coaching. Finally, we draw some conclusions.

2 Related Work

During recent years, topological relations have been much investigated in the static environments [29, 49]. Algebraic topological model for spatial objects was introduced in [49]. Thirteen topological relations between two temporal intervals were identified by [1]. After 4-intersection model [11, 12], the famous 9-intersection approach [13, 14] was proposed for formalism of topological relations. This approach is based on point-set topological concepts. Some drawbacks of such point-based topological approach is reported in [22]. The other significant approach known as RCC (Region-Connection Calculus) has been provided by [7, 8, 9, 22]. RCC as a pointless topology is based upon a single primitive contact relation, called connection, between regions. In this logic-based approach, the notion of a region as consisting of a set of points is not used at all. A similar method, so-called Mereotopology, is developed in [2, 47].

Due to problems of updating current location continuously or even per some seconds, such as bandwidth consumption, update transaction problem and query transaction problem [5, 6, 27], it is reasonable that instead of updating continuously, a predicting method will be used. The works [41, 26] and [39] offered a framework for modeling the movement of objects or individuals, processing of queries and multiple granularities. Assuming two known locations and a certain velocity are given, they modeled geospatial lifeline by lifeline bead. A lifeline bead consists of intersection of two inverted cones.

A method for reducing the size of computation is computation slice[20, 42]. The computation slicing as an extension of program slicing is useful to narrow the size of the program. It can be used as a tool in program debugging, testing, and software maintenance. Unlike a partitioning in space and time, which always exists, a distributed computation slice may not always exist [20].

Among others, two works using divide and conquer idea, called honeycomb and space-time grid, are closer to our proposal. The honeycomb model [15] focuses on temporal evolution of subdivisions of the map, called spatial partitions, and give a formal semantics for them. This model develops to deal with map and temporal map only. In [6, 7] the concept of space-time grid is introduced. Based upon the space-time grid, they developed a system to manage dynamically changing information. In the latter, they attempt to use the partitioning approach instead of an indexing one. This method can be used for storing and retrieving the future location of moving object.

In the previous work of the authors [32-38] applications of partitioning in space-time and using influenceability in motion planning, finding a collision-free path and relief management was demonstrated. This article can be considered as a theoretical extension of them.

3 Preliminaries

Causality is a well-known concept. There is much literature on causality, extending philosophy, physics, artificial intelligence, cognitive science and so on (e.g. [3, 30, 48]). In our view, influenceability stands for spatial causal relation, i.e. objects must come in contact with one another; cf. [3]. Although influenceability as a primary relation does not need to prove, it has some exclusive properties which show why it is selected. Influenceability supports contextual information and can be served as a basis for context aware mobile computing which has attracted researchers in recent years [17, 43]. This relation can play the role of any kind of accident and collision. It is well-known that the accident is the key parameter in most transportation systems (for example see [44]). As an example, the probability of collision defines the GPS navigation integrity requirement. In addition, the consideration of causal relation is closer to a naïve theory of motion [40].

In the relativistic physics [28, 46] based on the postulate that the vacuum velocity of light $\boldsymbol{c}$ is constant and maximum velocity, the light cone can be defined as a portion of space-time containing all locations that light signals could reach from a particular location (Figure 1). With respect to a given event, its light cone separates space-time into three parts, inside and on the future light cone, inside and on the past light cone, and elsewhere. An event ***A*** can influence (influenced by) another event; ***B***; only when ***B*** (***A***) lies in the light cone of ***A*** (***B***). In a similar way, the aforementioned model can be applied for the moving objects. Henceforth, a cone is describing an agent in mobile GIS environment for a fixed time interval. That means, a moving object is defined by a well-known acute cone model in space-time[24, 26]. This cone is formed of all possible locations that an individual could feasibly pass through or visit. The current location or apex vertex and speed of object is reported by navigational system or by prediction. The hyper surface of the cone becomes a base model for spatio-temporal relationships, and therefore enables analysis and further calculations in space-time. It also indicates fundamental topological and metric properties of space-time.

As described in Malek [34, 36], the movement modeling, are expressed in differential equation defined over a 4-dimensional space-time continuum. The assumption of a 4-dimensional continuum implies the existence of 4-dimensional spatio-temporal parts. It is assumable to consider a continuous movement on a differential manifold ***M*** which represents such parts in space and time. That means every point of it has a neighborhood homeomorphic to an open set in $\boldsymbol{R}^{\boldsymbol{n}}$. A path through ***M*** is the image of a continuous map from a real interval into ***M***. The homeomorphism at each point of ***M*** determines a Cartesian coordinate system (x_0, x_1, x_2, x_3) over the neighborhood. The coordinate x_0 is called time. In addition, we assume that the manifold ***M*** can be covered by a finite union of neighborhoods. Generally speaking, this axiom gives ability to extend coordinate system to the larger area. This area shall be interpreted as one cell or portion of space-time. The partitioning method is application dependent. The partitioning method depends on application purposes [5, 49] on the one hand, and limitation of the processor speed, storage capacity, bandwidth, and size of display screen [50] on the other.

4 Algebraic and Topological Structure

Let us take influenceability as an order relation (symbolized by π) and be primitive relation.

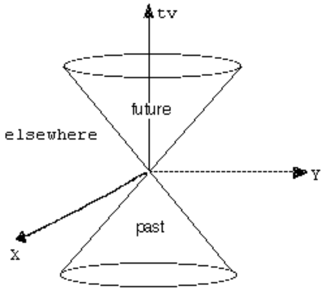


Fig. 1. A cone separates space-time into 3 zones, past, future, and elsewhere.

It is natural to postulate that influenceability is irreflexive, antisymmetric, but transitive, i.e.,

$$(x \text{ p } y) \wedge (y \text{ p } z) \Rightarrow x \text{ p } z$$

Thus, it can play the role of 'after'.

Definition 1 (Temporal order): Let x and y be two moving objects with t_x and t_y corresponding temporal orders, respectively. Then,

$$(x \text{ p } y) \Rightarrow (t_x < t_y)$$

Connection as a reflexive and symmetric relation [10]can be defined by influenceability as follows:

Definition 2 (Connect relation): Two moving objects x and y are connected if the following equation holds;

$$(\forall xy)C(x,y) := [(x \text{ p } y) \vee (y \text{ p } x)] \wedge \{\neg(\exists a)[(x \text{ p } a \text{ p } y) \vee (y \text{ p } a \text{ p } x)]\}$$

Consequently, all other exhaustive and pairwise disjoint relations in region connected calculus (RCC), i.e., *disconnection* (DC), *proper part* (PP), *externally connection* (EC), *identity* (EQ), *partially overlap* (PO), *tangential proper part* (TPP), *nontangential proper part* (NTPP), and the inverses of the last two; TPPi and NTPPi; can be defined.

Definition 3 (Immediately before): Let x and y be two moving objects with t_x and t_y corresponding temporal orders, respectively. Then, x is immediately before y (y is immediately after x) if:

$$t_x << t_y := [(t_x < t_y) \wedge (\neg \exists z(x \text{ p } z \text{ p } y)] .$$

The consensus task as an acceptance of the unique framework in mobile network can not be solved in a completely asynchronous system, but as indicated by Malek [34] with the help of influenceability and partitioning concept, it can be solved. Another task in mobile network is leader election. The leader, say *a*, can be elected by the following conditions:

$$\forall x \in \{The\ set\ of\ moving\ objects\} : a \prec x\,.$$

Furthermore, some other relations can be defined, such as which termed as *speed*-connection (SC) and time proper overlap (TPO) (see Figure 2):

$$SC(x,y) := \neg EQ(x,y) \wedge$$
$$\{[C(x,y) \wedge (\forall ab)\,(EC(x,a) \wedge (EC(x,b) \wedge EC(y,a) \wedge EC(y,b)] \Rightarrow C(a,b)\}$$
$$TPO(x,y) := \{(x \prec y) \wedge (PO(x,y) \wedge [\ \forall z\ (SC(x,z) \Rightarrow PO(y,z))]\ \}$$

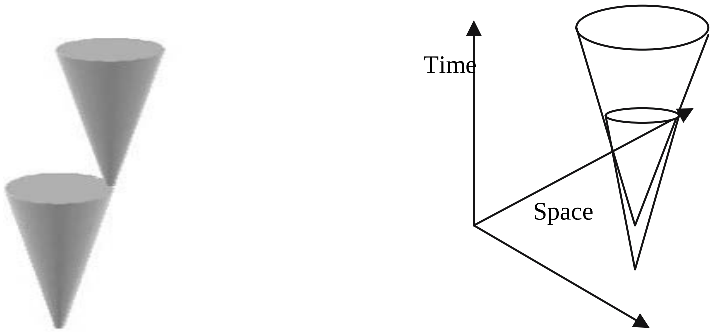


Fig. 2. a) Speed-connection relation and **b)** Time-proper relation between two objects

5 Expressivity Power

What has been shown so far is that if we regard a moving agent in mobile GIS environment as a cone then we can express certain important relations over agents purely in terms of the influenceability. In [36] we illustrated the expressive power of the theory. As an example four different relations between two objects; say A and B; can be discriminated by influenceability when A is before B and they are partial overlap. We are going to continue by some other relations and coaching problem, as well.

5.1 Qualitative Geometry

Influenceability can be served as a basis of affine geometry. We are not going into detail, however, but offer collinearity conditions which are very important for building a qualitative affine geometry.

Collinearity in time axis (COLT): Let x and y be two moving objects. They are collinear in time axis if

$$COLT(x,y) := \{EQ(x,y) \vee [\exists z\, ((TPO(x,z) \Rightarrow TPO(z,y)) \vee (TPO(y,z) \Rightarrow TPO(z,x))]\}.$$

Collinearity in space (COLS): Let x, y and z be three moving objects. They are collinear in space (Figure 3) if

$$COLS(x,y,z) := \{EQ(x,z) \vee EQ(y,z) \vee [\forall a((C(x,a) \wedge C(y,a)) \Rightarrow C(z,a))]\}.$$

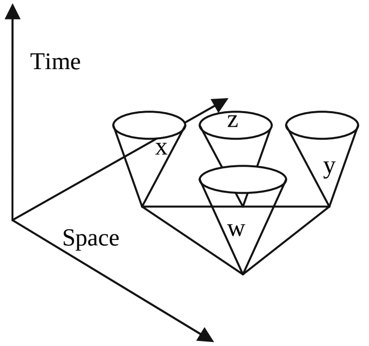


Fig. 3. x, y and z are collinear, but x, y and w are not.

5.2 Coaching

Team arrangement is an important task of any team coaching. Team arrangement in a mobile environment finds its applications not only in online problems such as Robocup (Robot World Cup) or battlefield problem, but also in offline coaching. The main assumptions about mobile environment are valid in the usual coaching problem. Robocup is a well known application area of this problem. It plays the role of a benchmark for many artificial intelligence and computation algorithms. In this scenario players can be modeled with a cone based on their estimated speed and position (Figure 4).

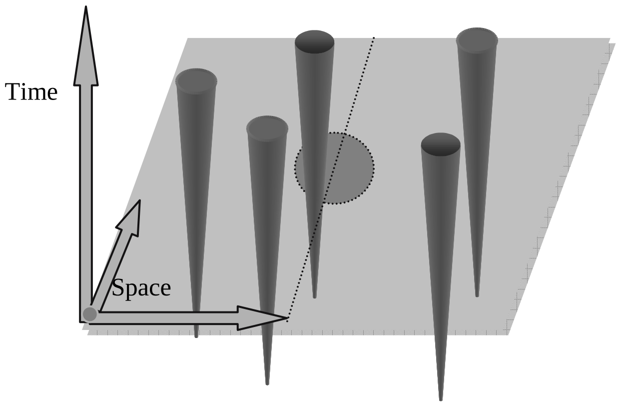


Fig. 4. Robocup soccer from influenceability view point

6 Conclusion and Further Work

There still exists shortcomings of a theoretical framework for qualitative spatial reasoning, especially for topological relationships in a mobile GIS environment. This paper has demonstrated that concerns to mobile GIS theory can profitably be addressed in terms of the partition and conquer idea, and influenceability relation. Of particular significance is the fact that the suggested idea can be given in a way that a context-aware and unique framework is prepared for mobile GIS environment. The expressive power of this idea is shown by some examples in section 4.

There are some possible directions for further work. Preparing a qualitative geometry based on influenceability is one of our future works. In addition, if it is possible to use a distributive lattice to describe influenceability, then, using Birkhoff's Representation Theorem [10] has significant computational advantages.

Acknowledgement

The first author is partially supported by Ministry of Science, Research, and Technology of Iran under grant 805935. Thanks to Dr. Winter (University of Melbourne) for his comments.

References

1. Allen, J.F.: "Maintaining Knowledge about Temporal Intervals", Communications of the ACM, 26(11), pp. 832-843, 1983.
2. Asher, N. and Vieu, L.: "Towards a geometry of common sense: a semantics and a complete axiomatisation of mereotopology", Proc. of *IJCAI95*, 1995.
3. Born, M.: "Natural Philosophy of Cause and Chance", Dover Publ., NY, 1949.
4. Caduff, D.: "Sketch-Based Queries in Mobile GIS-Environments", Master Thesis, Maine University, 2002.
5. Chon, H., Agrawal, D. and Abbadi, A. E.: "Storage and Retrieval of Moving Objects", Proceeding of International Conference on Mobile Data Management, 2001.
6. Chon, H., Agrawal, D. and Abbadi, A. E.: "Using Space-Time Grid for Efficient Management of Moving Objects", Proceeding of MobiDE, Santa Barbara, USA, 2001.
7. Chon, H., Agrawal, D. and Abbadi, A. E.: "Query Processing for Moving Objects with Space-Time Grid Storage Model", Technical Report No. 2001-15, Dept. of Computer Science, University of California, 2001.
8. Cohn, A., Bennet, B., Gooday, J. M. & Gotts, N.: "RCC: a calculus for region based qualitative spatial reasoning", GeoInformatica, 1, pp. 275-316, 1997.
9. Cohn, A. G. and Hazarika, S. M.: "Qualitative Spatial Representation and Reasoning:an Overview", Fundamenta Informaticae, 43, pp. 2-32, 2001.
10. Davey, B. A. and Priestly, H. A.: "Introduction to Lattices and Order", Cambridge University Press, UK, 1990.
11. Egenhofer, M.J.: "A Formal Definition of Binary Topological Relationships", Proceeding of Third International Conference on Foundations of Data Organization and Algorithms (FODO), Paris, France, Springer, 1989.
12. Egenhofer, M. J. and Franzosa, R. D.: "Point-set topological spatial relations", International Journal of GIS, 5(2), pp. 161-174, 1991.
13. Egenhofer, M. J. and Herring, J. R.: "Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases", Department of Surveying Engineering, University Maine, Orono, ME, 1991.
14. Egenhofer, M. J.:"A Model for Detailed Binary Topological Relationships", Geomatica, 47(3 & 4), pp. 261-273, 1993.
15. Erwig, M. and Schneider M.: "The Honeycomb Model of Spatio-temporal partitions", Proceeding of Spatio-Temporal database management, Edinburgh, Scotland, Springer, 1999.
16. Ellis, C. S.: "The case for higher-level power management", Proc. of 7th IEEE Workshop on Hot Topics in Operating System, Rio Rico, 1999.
17. Ferscha, A. and Hoertner, H. and Kotsis, G.: "Advances in Pervasive Computing", Austrian Computer Society, 2004.
18. Flinn, J. and M., S.: "Energy-aware Adaptation for Mobile Application", Proc. of Symposium on Operating Systems Principals, 1999.

19. Forman, G. H. and Zahorjan, J.: "The Challenges of Mobile Computing", IEEE Computer, 27(4), pp. 38-47, 1994.
20. Garg, V.K. and Mittal N.: "Computation Slicing: Techniques and Theory", Proceeding of DISC 2001, Lisbon, Portugal, 2001.
21. GIS-LOUNGE: "http://gislounge.com/ll/mobilegis.shtml", access Sept. 2005.
22. Gotts N.,Gooday J. & Cohn A.: "A connection based approach to common-sense topological description and reasoning", The Manifest, 79(1), 1995.
23. Heltzel, P.: "Wireless Computing and Networking", Techmedia, 2002.
24. Hendricks, M.D. and Egenhofer M. and Hornsby K.: "Structuring a wayfinder's Dynamic space-Time environment", Proceeding of COSIT 2003, Kartause Ittingen, Switzerland, Springer, 2003.
25. Hjelm, J.: "Creating Location Services for the Wireless Web: Professional Developer's Guide", John Wiley & Sons, 2002.
26. Hornsby, K. and Egenhofer, M. J.: "Modeling Moving Objects over Multiple Granularities", In: *Annals of Mathematics and Artificial Intelligence*, 2000.
27. Karimi, H. and Amin, H.: "Telegeoinformatics: Location-Based Computing and Services", CRC Press, 2004.
28. Kaufmann, W. J.: "Relativity and Cosmology", 1966.
29. Laurini, R. and D. Thompson, "Fundamentals of Spatial Information Systems", APIC Series, Academic Press, 1992.
30. Lewis, D.: "Causation", Journal of Philosophy, 70, pp. 556-567, 1973.
31. Li, L., Li, C. and Lin, Z.: "Investigation on the Concept Model of Mobile GIS", Proceeding of Symposium on Geospatial theory, Processing and Applications, Ottawa, 2002.
32. Malek, M.R.: "Motion Modeling in GIS", (in Persian) Proceeding of GEOMATIC 82, National Cartographic Center, Tehran, Iran, 2003.
33. Malek, M. R. and Delavar, M. R.: "Motion Modeling in GIS: A Case Study in Collision-Free Space-Time", URESCO, Modeling for Wayfinding Servieces, Germany, 2003.
34. Malek, M. R.: "A Logic-Based Framework for Qualitative Spatial Reasoning in Mobile GIS Environment with Applications to Navigations", In: Ferscha, A. and Hoertner, H. and Kotsi, G. (Ed.s):" Advances in Pervasive Computing", Austrian Computer Society, 2004.
35. Malek, M. R., Delavar, M. R. and Aliabady, S.: "The Location-Based services in a Mobile Government", (in Persian) Proceeding of The First Conference on IT, Software and E-city, Mashhad, 2004.
36. Malek, M. R.: "A Logic-Based Framework for Qualitative Spatial Reasoning in Mobile GIS Environment", Lecture Notes in Artificial Intelligence, 3066, pp. 418-426, 2004.
37. Malek, M. R. and Delavar M. R: "Mobile GIS: A Prototype for Mashhad City", (in Persian) Surveying Magazine, 16(1), pp.10-13, 2005.
38. Malek, M.R. and Delavar M.R: "A Mobile Spatial Information System for Rescue and Relief management", (in Persian) Proceeding of Second Conference of Scientific & Research on Relief and Rescue Management, Tehran, Iran, 2005.

39. Mark, D., Egenhofer, M., Bian, L., Hornsby, K., Rogerson, P. and Vena, J.: "Spatio-Temporal GIS Analysis for Environmental Health using Geospatial Lifelines", In Flahault, A., Toubiana, L. and Valleron, A.: 2nd International Workshop on Geography and Medicine: GEOMED'99, France, 1999.
40. McClosky, M.: "Naive theories of motion" In: Gentner D. and Stevens S. (Editors): "Mental Models", Hillsdale, New Jersey, Lawrence Erlbaum, 1983.
41. Miller, H. J., "Modelling accessedibility using space-time prism concepts within geographical information systems", International Journal of Geographical Information Systems, 5(3), pp. 287-301, 1991.
42. Mittal, N., Garg, V.K., "Computation Slicing: Technique and theory", Technical Report, TR-PDS-2001-002, University of Texas at Austin, 2001.
43. Nivala, A. M. and Sarjakoski, L. T.: "Need for Context-Aware Topographic Maps in Mobile Devices", In: Virrantaus, K. and Tveite, H.: *ScanGIS'2003*, Espoo, Finland, 2003.
44. Sang, J.: "Theory and Development of GPS Integrity Monitoring System", PhD Thesis, Queensland University of Technology, 1996.
45. Satyanarayanan, M.: "Fundamental Challenges in Mobile Computing", In:*ACM Symposium on Principles of Distributed Computing*, 1995.
46. Sexl, R. and Sexl, H.: "Weisse Zwerge- Schwarze Locher: Einführung in die relativistische Astrophysik", Trans. by Mansoory R., Sharif University, 2nd Edition, 1370.
47. Smith, B., "Mereotopology: A Theory of Parts and Boundaries", Data and Knowledge Engineering, 20, pp. 287-303, 1996.
48. Verma, T. S.: "Causal Networks: Semantics and expressiveness", In: Sacher R. and Levitt T.S. and Kanal L.N. (ed.s):"Uncertainty in Artificial Intelligence", Elsevier Science, 4, 1990.
49. White M. S., "A Survey of the Mathematics of Maps", Proceeding of Auto Carto IV, 1979.
50. Worboys M. F., "GIS: A Computing Perspective", Taylor & Francis, 1995.
51. Zhao, Y.: "Vehicle Location and Navigation Systems", ITS Series, Artech House, 1997.

From Turn-By-Turn Directions to Overview Information on the Way to Take

Kai-Florian Richter

Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, Universität Bremen, Germany
richter@sfbtr8.uni-bremen.de

Abstract

This chapter examines the in-advance route directions that provide a coarse overview on the way to take. We discuss the required properties of these directions in order for them being useful and present some first methods to generate such instructions. These methods are based on the model for context-specific route directions as presented in Richter & Klippel (2005); cardinal directions, global landmarks, and environmental structure, like different districts, are used as elements in our approach to generate coarse directions that indicate major reorientation points along the way. The generation process results in an abstract relational specification of these directions that reflect conceptual elements of route information; the abstract specification may then be externalized in different modalities, for example verbally or graphically, to be presented to the user, which will be outlined.

1 Introduction

One major application of today's mobile systems is wayfinding support. Nowadays, Car navigation systems are in wide use and more and more systems for bikers and pedestrians can be bought off the shelf. Even if wayfinding support is not the system's main purpose, such as is the case with location based services (LBS) - their aim is to provide a user with location-specific information on her current requests - usually there is still some underlying functionality that supports wayfinding: in the case of LBS, for example, to provide users with instructions to a requested shop or gas station in the vicinity of their current position.

We term such instructions for getting to a specific destination route directions (e.g., Denis, 1997). They are task-oriented specifications to be carried out to reach the destination (e.g., Tversky & Lee, 1998; Schweizer et al., 2000). We use the term route directions generically to refer to any kind of instructions for following a route; verbal, graphical, gestures, or a mixture of these. Route following comprises two basic processes: getting to a decision point and once there, determining the further direction to take (e.g., Daniel & Denis, 1998). That is, the main purpose of route directions is to support a wayfinder in deciding on how to proceed at a decision point.

Route directions can be distinguished into two broad categories: in-advance and incremental route directions. In-advance directions are presented to the user before the trip starts. They provide instructions on the complete route, i.e., on every decision point between origin and destination. This kind of route directions is, for example, generated by internet route planners. In incremental route directions, instructions are given step-wise, a single instruction for just the decision point the wayfinder is currently approaching. Such instructions are typically generated by mobile systems as here the device's location is assumed to be co-located with the user's which enables the device to determine the required timing for issuing the next instruction (cf. also Maaß, 1993; Habel, 2003a).

The latter - incremental route directions - are sufficient to keep users on track and allow them to keep their cognitive load low as they do not need to remember any instructions. However, they do not offer any survey information, i.e., users have no idea on what to expect along their trip. This forces them to rely completely on the system in their wayfinding (cf. Willis, 2005). To reduce this limiting dependency and to be able to cope for potential system's failures, users should be provided with an overview on the route to take before their trip starts - or in fact any time they feel like it. In the following, we present an approach to generating such descriptions. The approach is based on the model for context-specific route directions (Richter & Klippel, 2005) which originally is designed to produce complete in-advance route directions but can be extended to match our purposes.

The next section introduces the model for context-specific route directions, focusing especially on its underlying systematics of route direction elements. Section 3 then discusses benefits and properties of the aimed for overview information in more detail, while Section 4 presents an outline of principles and methods to generate such overview information.

2 Context-Specific Route Directions

In our research on route directions, we focus on people's conceptualization of routes and the actions necessary to (successfully) follow them. We define conceptualization to be the (process of forming a) mental representation of a route. A route is represented as a sequence of decision point / action pairs (cf. also Daniel & Denis, 1998). Hence, more precisely, conceptualization is (the process of forming a) mental representation of an (expected) decision point sequence with their accompanying actions. We have developed a model that aims at creating route directions supporting this conceptualization. The generated route directions are easy to process, i.e. they support forming and processing a representation of the corresponding route. This also eases route following as understanding a route direction is a prerequisite for using it (cf. Dale et al., 2003).

We coin the route directions generated by our model *context-specific route directions*. We use this term to emphasize that our model explicitly adapts the resulting route directions to the situation at hand, i.e., to the current action to take in the current surrounding environment. This reflects Dey and Abbowd's definition of context: “[...] any information that can be used to characterize the situation of an entity” (2000, p.3). For this adaptation, we need to account for the characteristics (the *structure*) of the environment in which route following takes place. The structure of an environment strongly influences the kind of instruction that can be given. The following structural aspects contribute to this influence: the embedding of the route in the spatial structure surrounding it, the structure of that route itself, path annotations, and landmarks that are visible along the route. Furthermore, different reference systems provide alternatives to describe necessary actions to follow a route (e.g., Tenbrink, 2005). An analysis of routes and route directions as well as the spatial knowledge required to determine and interpret them results in a systematics of elements that may be used in route directions (cf. Richter et al., 2004; Richter & Klippel, 2005). This systematics is summarized in Table 1. It is the basis for our generation process of context-specific route directions.

Table 1. Systematics of Route Direction Elements

Global References	Environmental Structure	Paths, Routes, and Landmarks
cardinal directions	edges	egocentric references
global landmarks	districts	landmarks at decision point
	slant	landmarks between decision points
		distant landmarks
		linear and area-like landmarks
		path annotations

In our model, route directions are represented as abstract, relational terms. They are a conceptual representation of the action to take at a decision point. For each element of the systematics, we define corresponding relational terms, which instantiate all possibilities of referring to the elements in route directions. As an example, consider a situation where the required action at the first decision point of a route is to take the left branch which is also marked by a sign leading to a train station. The instruction corresponding to the required action may be represented as (*DP1,left*) using egocentric references. As an alternative, the same action may be represented as (*DP1,follow/sign 'station'*) using the sign to the train station as path annotation.

To generate context-specific route directions, we need to choose from all possibilities to represent an action that best fits our aim of easing conceptualization of the route to take. That is, for each decision point along the route we choose an abstract instruction which externalization best eases its conceptualization. This choice may depend on the kind of instruction chosen for previous or following decision points. Accordingly, the generation of context-specific route directions is realized as an optimization process. Initially, for each decision point all possible instructions are generated, i.e., each description that unambiguously marks the action to take. Then, in the optimization step, from each decision point's set of possibilities the instruction that is best according to the optimization criterion is chosen (cf. Richter & Klippel, 2005, for a discussion of possible optimization criteria). In optimizing, the model exploits an important principle of conceptualizing routes and giving route directions: spatial chunking, the combination of several decision points into a single instruction as it, for example, becomes apparent in instructions like “turn left at the third intersection” (cf. Klippel et al., 2003).

The model has been designed to produce complete in-advance route directions covering all decision points. But it can be extended to the generation of coarse route directions that provide an (in-advance) overview on a route. This is further elaborated in Section 4. The next section discusses properties of such overview information.

3 Overview Information on the Way to Take

Route directions as discussed in the introduction offer information turn-by-turn, i.e., on how to proceed for every decision point along a route - be it that the information is presented all at once (in-advance) or step-wise (incremental). This information is needed to correctly execute route following, i.e. to get from origin to destination along a specified route. Such route directions are segmented at decision points. Each instruction covers one or several decision points and following an instruction a wayfinder always ends at a decision point (cf. Habel, 1988; Klippel et.al., 2003).

Overview information on the way to take, on the other hand, provides only coarse route directions. Such route directions are well suited for an initial, quick overview. They allow a wayfinder to get an idea on what to expect along the route. That is, they provide a supplement to incremental route directions as offered by mobile systems. They limit a user's (perceived) dependency on the device during wayfinding since they do not need to follow the device's instructions blindly any more. To account for restrictions that play a role in developing and using such applications, like small display size of mobile devices and users' limited cognitive capacity (cf. Wahlster et al., 2001), we need to take into account certain principles of generating coarse route directions. Even more than in ordinary route directions (cf. Denis, 1997), coarse route directions should not distract and bother users with unnecessary detail. Therefore, we concentrate on those points along a route which are crucial to keep the right (overall) direction. At these points, significant changes occur. They are the major reorientation points along a route.

Generating route directions based on major reorientation points allows for providing an overview on a route. But, it can also be used for wayfinding assistance in partially familiar environments (cf. Schmid & Richter, 2006; Tomko & Winter, 2006a). In such environments where some areas and major routes are known to the wayfinder, detailed turn-by-turn directions are not necessary for these known parts. Instead, instructions like “go to the main station, I'll guide you from there” suffice in this case (Schmid & Richter, 2006). Tomko and Winter (2006a,b) propose an approach for such route directions. As initial element in route directions they select an element - here, a district - of an environment on the coarsest possible level of granularity. This reference gets more and more refined the closer the destination gets. The approach by Tomko & Winter is top-down or destination-based. The approach presented in this chapter is bottom-up or route-based. It focuses on abstracting route directions starting from turn-by-turn instructions and can be seen to be a counterpart of Tomko and Winter's approach.

Concentrating on major reorientation points corresponds to the planning level in wayfinding as explained, for example, in Timpf et al. (1992). While the level of actions requires information on all decision points in order to take the correct turn, the planning level requires less granularity in information, i.e., less detail. On this level, coarse information is sufficient as the aim is to provide just an overview without bothering users with details on how to actually execute route following. This is, for example, reflected in the approach by Höök (1991), who generates route directions for local residents who are assumed to know the place fairly well. In her approach, several roads are subsumed into a high-level instruction and details, like small roads, are omitted. Hence, conceptualization of such coarse route directions can also be only coarse and leaves many parts of the route underdetermined.

Coarse route directions do not guarantee that a wayfinder strictly follows the intended route, i.e., the route determined by the computational system. This is because segmentation of a route is not done at decision points, but is based on major reorientation points. These points divide a route into regions. The regions comprise of the area between two reorientation points; each instruction in a coarse route direction covers one such region. Coarse route directions guide a wayfinder from one region to another without fixing a specific route between these regions. Consequently, if just relying on coarse directions, it is up to the wayfinder to fill these gaps with her own decisions on the exact route to take (for an overview on human region-based navigation see Wiener & Mallot, 2003). In case of combining coarse route directions with incremental route directions, the mobile device may provide information for these in-between routes.

4 Generating Overview Information: An Outline

In order to generate overview information on a route, i.e., coarse route directions, the major reorientation points along the route and their accompanying regions need to be identified and instructions providing coarse information on how to reach these points need to be generated. To that end, we make use of the elements of the systematics (see Table 1) that are applicable in coarse route directions. Looking at this systematics, those elements on coarser levels of granularity, i.e., those that abstract from single decision point / action pairs to a great extent are suited for generating overview information. This holds especially for elements of the first two levels of the systematics - the level of global references and the level of environmental structure. From the level of paths, routes, and landmarks the elements distant landmark, linear and area-like landmarks, and path annotation are used as they also strongly abstract from single decision point / action pairs.

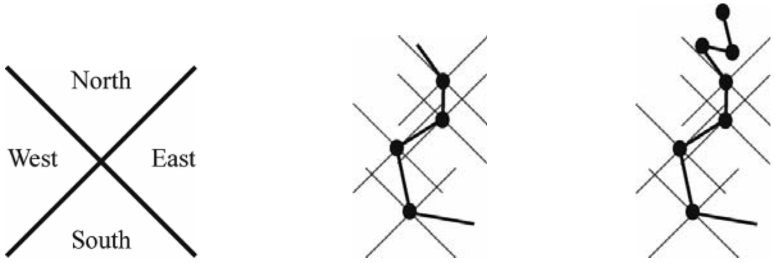
Except for cardinal directions, instructions using these elements need to include a statement announcing until which point they hold, i.e., when the corresponding action like following a linear landmark ends and a change of action occurs. We term such a statement end qualifier; an example for an end qualifier is “until the gas station” in an instruction like “follow the river until the gas station”. End qualifiers are required with those systematics' elements that allow combining (potentially) many decision points into a single instruction (cf. Klippel et al., 2003; Richter & Klippel, 2005). They announce changes of action after a (potentially) considerable part of a route. Hence, end qualifiers play an important role in coarse route directions. As argued in the last section, segmentation of coarse route directions is done along major reorientation points, not at decision points. End qualifiers are well suited to mark these reorientation points. In the same line, confirmation information, which is used in detailed route directions to assure a wayfinder that she is still on the right track, may become ‘real' instructions in coarse route directions. Coarse directions are supposed to indicate the overall direction towards the destination. Confirmation information, like “you will cross the river”, are well suited to indicate this direction - since that is exactly what they are used for in detailed route directions.

Consequently, for generating coarse route directions there are two elements of our representation of routes that may be exploited: first, those decision points that mark the end of a chunk, i.e. decision points at which an environment's feature, usually a landmark, may be exploited as an end qualifier for instructions. Second, route segments along which confirmation information can be determined may be part of a representation of coarse route directions. Confirmation information can be based on references to edges (“cross edge”) or landmarks between decision points (“pass landmark”). These route elements mark the border of a region which a wayfinder may pass without (significant) change of action; termed *region of equal directedness*. This directedness is equal relative to some feature, i.e., to one of the route direction's elements. Examples include “follow the river” or “go in direction of the TV tower”. In the former example, a linear landmark induces the directedness - ‘keep next to the river' - in the latter, a distant landmark sets the direction - ‘lessen your distance to the tower and keep it in front of you'.

Such directions typically hold for several consecutive decision points. Each such sub-sequence of a route's decision points forms one of the regions induced by major reorientation points. Now, to fulfill the requirements of overview information discussed in Section 3, i.e. providing an initial idea on the route without distracting a user with a lot of details, the aim in generating coarse route directions is to cover as much of the route with as few reorientation points as possible. That is, coarse route directions should comprise of a few, large regions. Like in the original model for context-specific route directions (see Section 2), generating overview information can be solved as an optimization problem. The optimization process is similar to the original one. Accordingly, we can re-use the same algorithms.

Determining coarse route directions requires some heuristics. First of all, if we aim for as few reorientation points as possible, we, consequently, aim for as few chunks as possible. In other words, we are looking for chunks that cover as much decision points as possible. The optimization criterion is to aim for the minimal number of chunks. For optimization purposes, we need to determine the chunks using the systematics' elements suitable for coarse route directions. Just as with context-specific route directions, we start by creating every possible chunk and use this set of chunks as basis for optimization. This way, it is guaranteed that we cover the complete route. However, this may not be desired as it may lead to more detail in the coarse directions than necessary. Therefore, we need to apply additional heuristics which on the one hand allow leaving parts of the route unconsidered, and on the other hand coarsening instructions by abandoning the need to generate directions that are necessarily unambiguous.

Development of these heuristics is current work. They all work in a similar way: they add decision points to the region of equal directedness as long as the element defining this region is applicable. As an example, let us consider generating coarse direction information using cardinal directions. First, to determine such directions, we need a cardinal direction model like, for example, one of those presented by Frank (1992). This model allows calculating the cardinal direction to take (e.g. 'north', 'southeast', etc.) at each decision point. For our purposes of coarse directions we choose a four-sector model ('north', 'east', 'south', 'west') that itself already provides just coarse information (see Fig. 1a). A possible heuristic is to add decision points to the cardinal direction-chunk as long as the decision point at hand lies in the previous sector and the direction to take corresponds to the initial direction (Fig. 1b).



a) b) c)

Fig. 1: a) four-sector model of cardinal directions (cf. Frank, 1992); b) chunking decision points (the dots) of a route (the dashed line) with equal directedness based on the four-sector model; c) small deviations from the overall direction need to be ignored.

As an open issue remains the question how we can deal with minor deviations from the overall direction? That is, we need to extend the heuristics in such a way that in determining coarse cardinal directions it ignores small route-segments that lead in different directions (see Fig. 1c). To that end, two factors may be used for a threshold: the length of the deviating route-segment and the degree of deviation, i.e. the deviating angle's size. This is a recurring pattern for all the heuristics. For all elements, minor deviations, i.e., small parts of the route where an element is not applicable need to be ignored which requires thresholds that allow calculating whether a decision point is to be added to the region of equal directedness even though the element is not (unambiguously) applicable here.

Another open issue is the externalization of coarse route directions, i.e. ways to present this in-formation to users. Verbal presentation - either written or spoken - is easily realizable by developing a parser for the abstract directions generated by our model and seems to be well suited, since verbal instructions typically are underdetermined and may leave many relations unspecified. Graphical presentation, on the other hand, needs to settle for exactly one instantiation due to the representation medium's properties (Habel, 2003b). It is, therefore, often taken to represent veridical information. Hence, suitable schematization means need to be developed to indicate that the information presented is only coarse and may not be complete (cf. Agrawala & Stolte, 2001; Klippel et al., 2005, for such approaches).

5 Summary

In this chapter, we presented ongoing work on an approach for providing overview information on the way to take. It is based on our model for context-specific route directions which allows generating in-advance route directions that aim at being easily conceptualizable. This model can be extended to determine coarse route directions which concentrate on the major reorientation points along a route. We outlined how to determine these reorientation points based on an optimization process and suggested some initial heuristics for further abstractions needed to concentrate just on the crucial information.

Coarse route directions provide an initial overview on the route to take, i.e., allow a wayfinder to get an idea of what to expect along a route without bothering and distracting her with unnecessary detail. In mobile systems providing incremental route directions a user is forced to rely on the system in her wayfinding. Overview information relieves a user from this (perceived) dependency as she does not need to follow the instructions blindly anymore. Hence, we argue that such coarse route directions ideally supplement incremental route directions as provided by mobile systems and should be incorporated in such systems as an option a user can choose.

Future work comprises development of additional abstraction heuristics, (graphical) externalization means for coarse route directions, and an evaluation of the model's performance.

Acknowledgements

This research is carried out as part of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition. Funding by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. I would like to thank Katharine Willis for commenting an earlier version of this chapter and Sabine Timpf for discussing ideas of this work.

References

Agrawala, M. & Stolte, C. (2001). Rendering effective route maps: Improving usability through generalization. In SIGGRAPH 2001, Los Angeles, CA.

Dale, R., Geldof, S., and Prost, J.-P. (2003). CORAL: Using natural language generation for navigational assistance. In Proceedings of the 26th Australasian Computer Science Conference (ACSSC2003). Adelaide, Australia.

Daniel, M.-P. & Denis, M. (1998). Spatial descriptions as navigational aids: A cognitive analysis of route directions. Kognitionswissenschaft, 7:45–52.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. In Cahiers Psych. Cognitive 16(4):409-458.

Dey, A.K. & Abowd, G.D. (2000). Towards a better understanding of context and context-awareness. In Proceedings of the Workshop on the What, Who, Where, When, and How of Context-Awareness. New York.

Frank, A.U. (1992). Qualitative spatial reasoning about distances and directions in geographic space. Journal of Visual Languages and Computing 3:343-371.

Habel, C. (1988). Prozedurale Aspekte der Wegplanung und Wegbeschreibung. In H. Schnelle and G. Rickheit (eds.), Sprache in Mensch und Computer (pp. 107-133). Opladen: Westdeutscher Verlag.

Habel, C. (2003a). Incremental generation of multimodal route instructions. In Natural Language Generation in Spoken and Written Dialogue, AAAI Spring Symposium 2003. Palo Alto, CA.

Habel, C. (2003b). Representational commitment in maps. In M. Duckham, M. Goodchild, and M. Worboys (eds.), Foundations of GIS. Taylor and Francis.

Höök, K. (1991). An Approach to a Route Guidance Interface. Licentiate thesis, Dept. of Computer and System Sciences, Stockholm University.

Klippel, A., Tappe, H., and Habel, C. (2003). Pictorial representations of routes: Chunking route segments during comprehension. In C. Freksa, W. Brauer, C. Habel, and K.F. Wender (eds.), Spatial Cognition III. (pp. 11-33). Berlin: Springer.

Klippel, A., Richter, K.-F., Barkowsky, T., and Freksa, C. (2005). The cognitive reality of schematic maps. In L. Meng, A. Zipf, and T. Reichenbacher (eds.), Map-Based Mobile Services - Theories, Methods, and Implementations (pp. 57-74). Berlin: Springer.

Maaß, W. (1993). A cognitive model for the process of multimodal, incremental route descriptions. In A.U. Frank & I. Campari (eds.), Spatial Information Theory: A Theoretical Basis for GIS (pp. 1-13). Berlin: Springer.

Richter, K.-F. & Klippel, A. (2005). A model for context-specific route directions. In C. Freksa, M. Knauff, B. Krieg-Brückner, & et al. (eds.), Spatial Cognition IV. (pp. 58–78). Berlin: Springer.

Richter, K.-F., Klippel, A, and Freksa, C. (2004). Shortest, fastest, - but what next? A different approach to route directions. In M. Raubal, A. Sliwinski, and W. Kuhn (eds.), Geoinformation und Mobilität, Beiträge zu den Münsteraner GI-Tagen (pp. 205-217). Münster: IfGI prints.

Schmid, F. & Richter, K.-F. (2006). Extracting places from location data streams. In UbiGIS 2006, Workshop at GIScience 2006. Münster, Germany.

Schweizer, K., Katz, S., and Janzen, G. (2000). Orientierung im Raum – kognitive Grundlagen und sprachliche Realisierung. Tourismus Journal 4(1):79-101.

Tenbrink, T. (2005). Semantics and application of spatial dimensional terms in English and German. Tech. Rep. No. 004-03/2005. SFB/TR 8 Spatial Cognition, Universität Bremen / Universität Freiburg.

Timpf, S., Volta, G. S., Pollock, D. W., Frank, A. U., & Egenhofer, M. J. (1992). A conceptual model of wayfinding using multiple levels of abstraction. In A.

Frank, I. Campari, & U. Formentini (eds.), Proc. of Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, (pp. 348–367). Springer.

Tomko, M. & Winter, S. (2006a). Identification of the initial entity in granular route directions. In A. Riedl, W. Kainz, and G. A. Endes (eds.), 12th International Symposium on Spatial Data Handling. Berlin: Springer.

Tomko, M. & Winter, S. (2006b). Recursive construction of granular route directions. In Journal of Spatial Science 51(1):101-115.

Tversky, B. & Lee, P.U. (1998). How space structures language. In C. Freksa, C. Habel, K.F. Wender (eds.), Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge (pp. 157-175). Springer.

Wahlster, W., Baus, J., Kray, C., & Krüger, A. (2001). REAL: Ein ressourcenadaptierendes mobiles Navigationssystem. Informatik Forschung und Entwicklung, 16:233–241.

Wiener, J.M & Mallot, H.A. (2003). ‘Fine to coarse' route planning and navigation in regionalized environments. Spatial Cognition a. Computation 3(4):331-358.

Willis, K. S. (2005). Situated cognition in the use of mobile devices for wayfinding tasks. In Workshop at Mobile HCI 2005, Salzburg, Austria.

Geocoding Japanese Walking Directions using Sidewalk Network Databases

Masatoshi Arikawa, Kouzou Noaki

Center for Spatial Information Science, University of Tokyo
arikawa@csis.u-tokyo.ac.jp

Abstract

This paper proposes an advanced method of geocoding for walking directions in documents using daily local expressions. Walking directions are usually described in casual expressions, but not in a formal address. Thus, it has been considered practically impossible to geocode walking directions. However, sidewalk network databases in major cities of Japan have been available for pedestrian navigation services with GPS-equipped mobile phones since 2003. The databases can be expected to enable more advanced geocoding for larger scale natural language expressions. We first introduce the core schema of sidewalk network databases. Then, we explain a structure of walking directions expressed in Japanese, and propose FRS (Formal Route Statement) to represent and process walking directions by means of a computer. Finally, prototype systems which have been developed based on our proposed framework are presented.

1 Introduction

Japan is one of the most advanced countries in the use of consumer-oriented digital map systems and services. For example, more than one hundred million car navigation systems have been sold in Japan. The car navigation systems provide voice navigation, urban three-dimensional computer graphic views and audio-visual systems such as hi-fi stereo DVD/HD/MP3 players. For persons who once used Japanese car navigation systems, it may be difficult to drive without them because of their high quality of service. In addition to car navigation systems, approximately twenty million cars have access to the real-time information, called VICS (Vehicle Information and Communication System) [1], such as traffic jam, accident, construction, parking and restaurants through FM radio, beacon, mobile phone and Internet in Japan. Furthermore, more than twenty million mobile phones are equipped with GPS in Japan. Japanese mobile phones are very different from those of western countries, and may be considered as a type of PDA that has evolved independently in Japan. Specification of up-to-date Japanese mobile phones includes full-color, high resolution display (240x320 pixel, 167dpi), 3D graphics engine, Java/flash/SVG compatibility, full web browser, full Internet mailer, diary, scheduler, video recorder, 2D barcode (QR code) reader, Bluetooth, IC card for electronic money, digital TV receiver, MP3 player, removable memory, GPS, digital compass and so on. It is popular for Japanese people to use maps and pedestrian navigation services on their mobile phones with integrated GPS and digital compass. The voice and textual navigation services on both car and pedestrian navigations are realized by using algorithms that both find the shortest path and convert a path to natural language sentences as well as by using databases of networks including information about roads, landmarks, sidewalks, crossroads, entrances of buildings and so on. Thus, detailed road networks covering all of Japan are available. Also, sidewalk networks for pedestrian navigation services are available in most major cities. We, therefore, have high quality and sufficient geographic network databases in Japan.

We have studied methods of converting Japanese geo-referenced descriptions like addresses and place names into their corresponding geographic coordinates [2]. The process of converting descriptions into geographic coordinates is called geocoding. In this paper, we focus on walking directions in Japanese which are often written in web pages of restaurants and shops, as a new type of target to geocode using rich Japanese sidewalk network databases.

2 Sidewalk Network Databases

Sidewalk network databases store data about underground walks, footbridges and cross walks for pedestrians. Sidewalk network databases are simply structured as nodes and links. A node has geometric coordinates. A link is defined as a vector between two nodes. Sidewalk network databases are provided as commercial products by Shobunsha Publications Inc. [3]. The commercial sidewalk network databases presently cover major cities in Japan.

Before the emergence of sidewalk network databases, there have been popular geographic network databases such as road networks for car navigations, railroads, facilities networks and so on. However, the previous geographic network databases were all designed for small scale uses, not for large scale uses such as pedestrian navigations. On the other hand, sidewalk network databases for walkers are getting popular for pedestrian navigation systems since 2003 in Japan. Some services and products using sidewalk network database have already been on the market. “EZnaviwalk” [4] provided by KDDI is one of the most popular pedestrian navigation services using a cell phone, GPS and electronic compass (Fig.2.). Using sidewalk network databases along with train timetables and airline timetables, EZnaviwalk finds the most direct, time-saving or money-saving route. This service was begun in October 2003.

Fig. 1. A view of the most popular pedestrian navigation service “EZnaviwalk” [4] on a mobile phone in Japan provided by KDDI using GPS and electronic compass.

The sidewalk network database used in our research is simply structured as nodes and links. A node has not only geometric attributes like coordinates but also non-geometrical attributes like name of spatial entity, its class and related url. A link has a distance and an angle from its start node to its end node. Users are allowed to extend the integrated database by entering additional nodes and links or inputting text data. In particular, “intersection”, “street” or “slope street” is shaped with the multiple nodes and links. For example, at least four nodes are needed to shape an intersection (Fig.2). Figure 3 shows the core schema of sidewalk network databases. A node has attributes, that is, *id* (the identifier of a node), *name* (a spatial entity's name which corresponds to the node), *coordinates* (tuples of longitude and latitude), *in* (a name of a street or an area both of which are composed of multiple nodes and links), *class* (a class of spatial object), *incoming link* (*id* of an incoming link), *outgoing link* (*id* of an outgoing link) and *poi* (additional information concerning the point of interest except for *name*, *class* and *url*)". A link has its *id* (the identifier of the link), *start_node* (*id* of the node from which the link starts), *end_node* (*id* of the node at which the link arrives), *direction* (of the link in degree) and *distance* (of the link in meter).

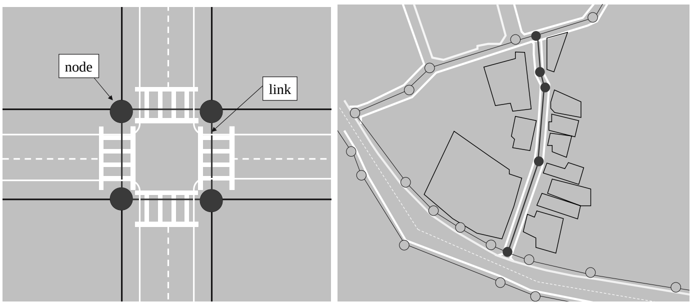


Fig. 2. Examples of nodes and links stored in sidewalk network databases. The left figure shows an intersection and the right one shows a slope street.

Sidewalk network DB :	(*Nodes*, *Links*)
Nodes :	A set of *node*
Links :	A set of *link*
node :	(*id*, *name*, *coordinates*, *in*, *class*, *incoming_link*, *outgoing_link*, *poi*)
link :	(*id*, *start_node*, *end_node*, *direction*, *distance*)
node.id :	*id* of the node
Node.name :	*name* of the node
node.coordinate :	*coordinates* (*longitude*, *latitude*)
node.in :	*name* of group
node.class :	*class* of spatial object
node.incoming_link	*id* of the incoming link
Node.outgoing_link :	*id* of the outgoing link
node.poi :	information for the point of interest
link.id :	*id* of the link
link.start_node :	*id* of the start node
link.end_node :	*id* of the end node
link.direction :	its direction in degree
link.distance :	its distance in meter

Fig. 3. Core schema of the sidewalk network databases used in our research

3 Natural and Formal Route Descriptions

3.1. Structure of Walking Directions

A walking direction can be divided into noun phrases and verbal phrases. For example, the description
“渋谷駅ハチ公口を出て道玄坂を上り交差点を右へ曲がる (You exit from the Shibuya Hachiko Exit and go up Dogen-zaka Slope Street, and then turn right at the intersection)” is divided as follows:

渋谷駅ハチ公口を出て道玄坂を上り交差点を右へ曲がる
↓
[渋谷駅ハチ公口][を出て][道玄坂][を上り][交差点][を右へ曲がる]
↓
noun phrases: [渋谷駅ハチ公口],[道玄坂],[交差点]
verbal phrases: [を出て],[を上り],[を右へ曲がる]

All of the above noun phrases have a place name. A place name refers to the unique place in the real world and works as a reference point for the next verbal phrase. The above verbal phrase has a preposition such as “を(at, from or through)”, a verb such as “曲がる(turn)” or “出る(exit)” and an adjunct such as “右へ(right)”, “100メートル(100meters)” or “5分(5minutes)”. It is usual for Japanese people that the agent noun for the verbal phrase is often omitted in a description. A person who is referred by an omitted agent noun is a walker who is a writer of a description or third person.

We call a noun phrase in a walking direction *spatial anchor description*. Spatial anchor descriptions includes unique noun and general noun. A general noun such as “intersection” is deduced from context of a description and situation. Spatial anchor descriptions referring to those places can be used as start points, passage points and end points of routes in walking directions. These verbal phrases are the expressions of a walker’s action on the route. Walker’s actions on the walking direction denote topology between places. We call a verbal phrase in a walking direction *spatial relationship description*.

3.2 Simple Model of Walking Directions

We simplify walking directions, such as a list structure in Figure 4. A description of spatial anchor corresponds to an ellipse and spatial relationship description corresponds to a rectangle in Figure 4. Figure 5 illustrates the image of geocoding a walking direction into spatial database. Constructing a walking direction can be compared to sticking of pins and stretching of a thread between pins on a map. Using this simple model, we define the formal statement of walking directions.

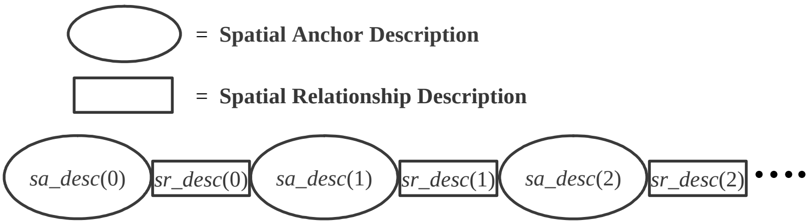


Fig. 4. Image of model for walking directions

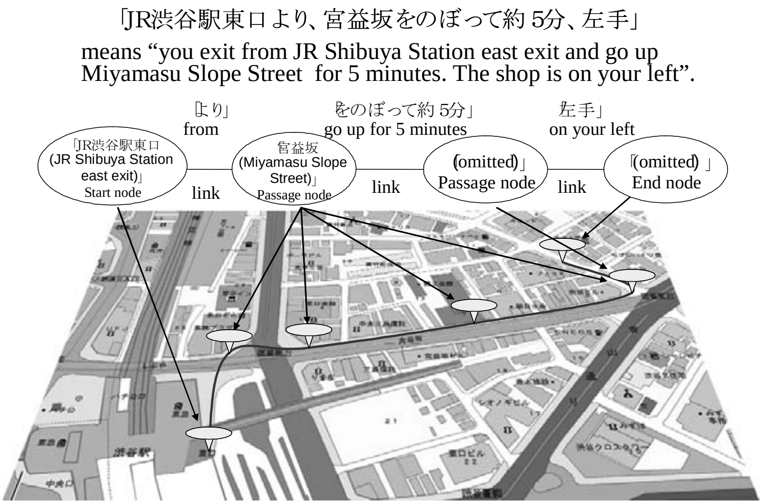


Fig. 5. Image of geocoding a walking direction with a spatial database through the intermediary of a simple model.

3.3 Grammar of Formal Route Statement (FRS)

Formal statements are necessary for computers to indirectly deal with walking directions. On the assumption that all of walking directions can be expressed with nodes and links, we propose *Formal Route Statement (FRS)* to represent and process walking directions. FRS also works as a query language for sidewalk network databases. Using FRS, a walking direction is represented as a sub-graph of a directed graph of sidewalk networks (Figure 6).

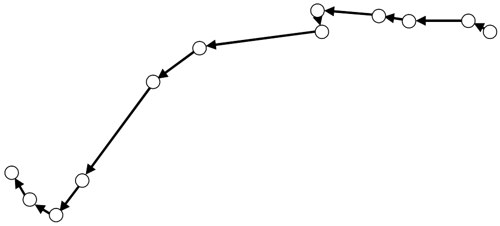


Fig. 6. Sub-graph *G* which is the result of FRS(*route_desc*), where *route_desc* is a text string of a walking direction.

Figure 7 shows the grammar of FRS. Generalization tables are indispensable for converting various casual descriptions into regular ones, one of which is FRS (Table 1). A use case of the generalization tables is to make an instance of the spatial relationship as a value of the attribute "*link.connect*" in Figure 7. The attribute "*link.connect*" plays an important role to find a spatial object when a name referring to the next place is omitted.

FRS ::=	sa_desc(0)(:sr_desc(i):sa_desc(i+1))* [i={0,...,n}];
sa_desc(i) ::=	node(i).node_attribute_list
node_attribute_list ::=	none | node_attribute_value (&node_attribute_value)*
Node_attibute_value ::=	node_attribute = value
node_attribute ::=	id | name | coordinate | class | status
value ::=	numerical_value | stirng_value | url | status_values | connect_values
status_values ::=	start | end | via
connect_values::=	straight | right | left
sr_desc(i) ::=	link(i).link_attribute_list
link_attribute_list ::=	none | link_attribute_value (&link_attribute_value)*
link_attibute_value ::=	link_attribute = value
link_attribute ::=	id | start_node(id) | end_node(id) | direction | connect | distance

Fig. 7. Grammar of a Formal Route Statement.

Table 1. A generalization table for descriptions of spatial relationship and values of the attribute "*link.connect*"

Specialized descriptions for spatial relationships	Generalized descriptions for spatial relationship (values of link.connect)
go forward, go ahead, advance	Straight
turn to the right, on the right, on one's right	Right
turn to the left, on the left, on one's left	Left

4 Prototype of Sidewalk Network Database Management System

We have developed a naive management system for sidewalk network databases. We explain each component in the user interface (Figure 8) as follows.

1. Menu button: Users can change the operation mode by the menu buttons. Main functions are (1) loading and saving sidewalk network data which are XML formatted and (2) adding, erasing and moving both nodes and links. Furthermore, we select functions of referring to node information (e.g., a name of a place) and filling it using the entry form.
2. Map area: Sidewalk network databases and a map image in the selected area are overlapped and visualized.
3. Display of node information: This area allows users to see values of attributes, a name, a class and a picture for an instance of a spatial object in the map area.

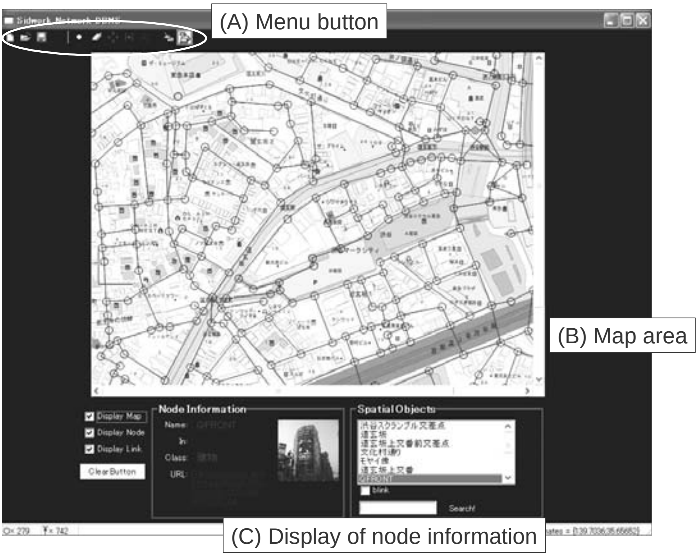


Fig. 8. Graphical user interface of the prototype system for managing sidewalk networks.

5 Prototype of Geoparser for Walking Directions

We have developed a prototype system, which processes a walking direction in Japanese, and then visualizes it as a polyline on the map using a sidewalk network database. Each component in the user interface (Figure 9) is as follows:

A. Input text form: Users can input a walking direction using this text form.
B. Output text form: A result of separating input text and validating separated elements is displayed in this form.
C. Output map area: A route is visualized on the sidewalk network database as a result of geocoding a walking direction.

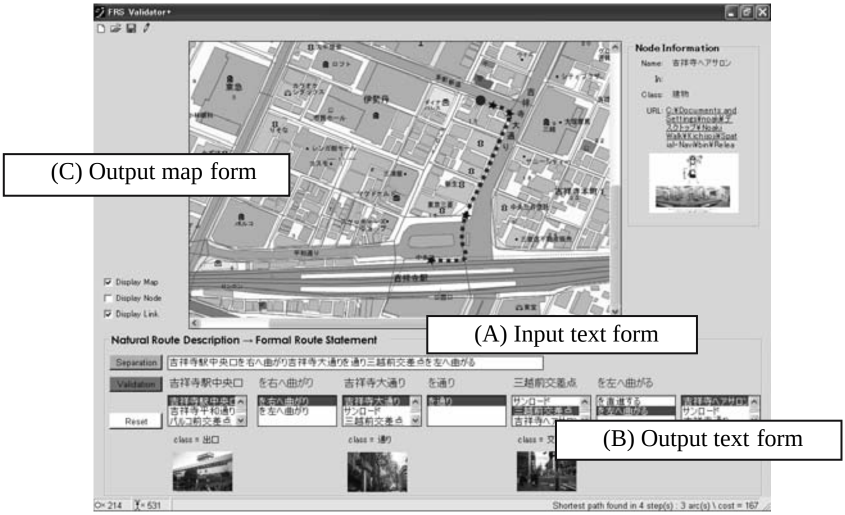


Fig. 9. Graphical user interface of the prototype system for geocoding walking directions.

Using a walking direction as an example of an input text, Figure 10 shows a behavior of the geocoding process for an example of a walking direction. The captions of the figures explain the details of the behavior.

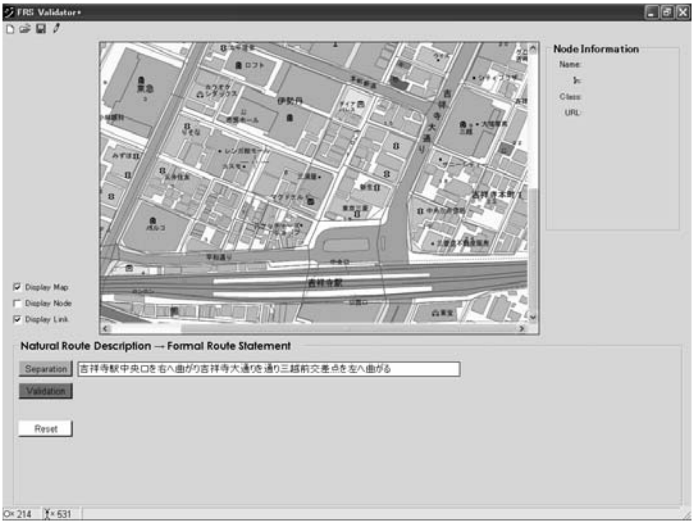


Fig. 10a. A user cuts a Japanese route description from a web page of a hair salon, and pastes it in the input text form of the prototype system.

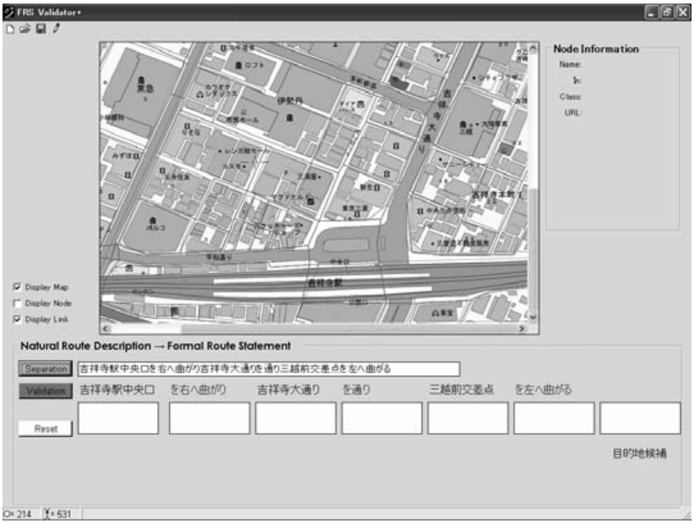


Fig. 10b. The user pushes the Separation Button, and then the system separates the route description into noun and verbal phrases using a morphological analysis method. We have adopted *Chasen* [5] as Japanese morphological analysis system in a component of our prototype system.

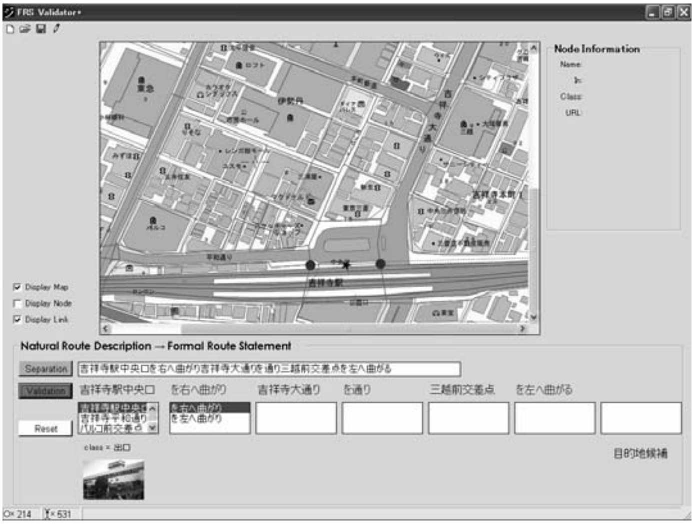


Fig. 10c. The user pushes the Validation Button, then the system tries to match the first noun phrase *sr_desc*(0) with a node of the sidewalk network database. The blue star symbol is the node matched with the first noun. Two red symbols are the next candidate nodes connected to the matched node.

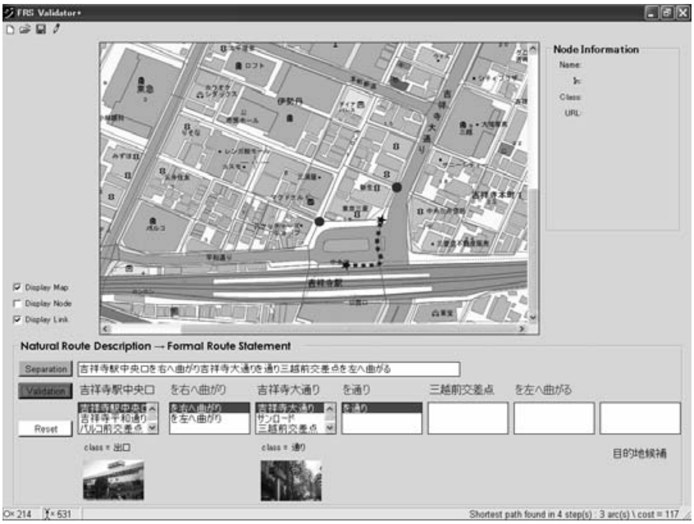


Fig. 10d. The user pushes the Validation Button, then the system tries to match the second noun phrase *sa_desc*(1) with a node of the sidewalk network database. Also, it searches the shortest path (the red dotted line) from the node matched *sa_desc*(0) to the nearest one of the nodes making up *sa_desc*(1). This figure shows the result of processing *sa_desc*(0) + *sr_desc*(0) + *sa_desc*(1).

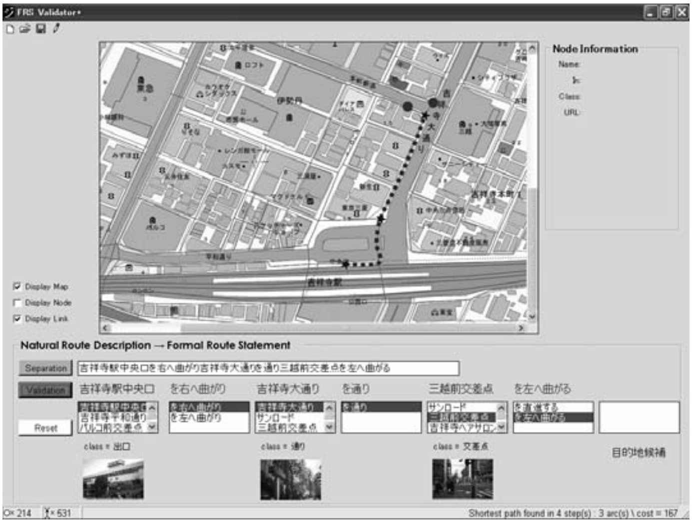


Fig. 10e. Pushing the Validation Button allows the system to search the shortest path to the nearest one of nodes matched with the third noun phrase *sa_desc*(2). This figure shows the result of processing *sa_desc*(0) + *sr_desc*(0) + *sa_desc*(1) + *sr_desc*(1) + *sa_desc*(2).

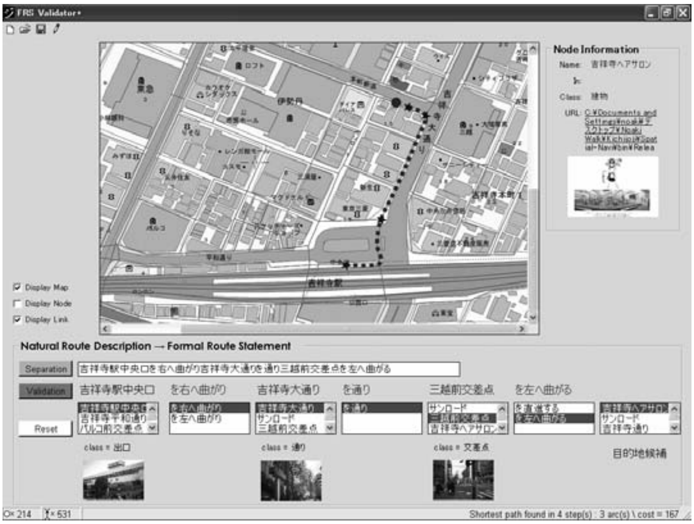


Fig. 10f. The end node is deduced from the description of *sr_desc*(2) and the direction of the last matched link. This figure shows the result of processing *sa_desc*(0) + *sr_desc*(0) + *sa_desc*(1) + *sr_desc*(1) + *sa_desc*(2) + *sr_desc*(2).

6 Conclusion

This study proposed an advanced method for geocoding geo-referenced descriptions. The establishment of these methods will make it easier to obtain local information using the spatial representation among ever increasing amount of spatial data. This paper showed a basic framework to geocode walking directions for pedestrians walking through a city by means of sidewalk network databases. On the basis of the structure of walking directions, we clarified the three significant ideas in our proposed method as follows:

- Formal Route Statement
- Schema of extended sidewalk network databases
- A method to validate Formal Route Statement

We aim for the realization of the robust system. For example, the system corrects route descriptions when invalid route descriptions are detected. In addition, there is a problem of ambiguity in natural language. In order to solve this problem, multiple solutions, which are possible routes to a destination, should be ranked by some criteria of quality of the geocoding results.

Acknowledgements

This research was supported in part by Grant-in-aid for Scientific Research on Priority Area “Informatics Studies for the Foundation of IT Evolution” by the Ministry of Education, Culture, Sports, Science and Technology of Japan. We would like to thank Dr. Takeshi Sagara and Dr. Hideyuki Fujita for their useful advice and comments.

References

[1] VICS, http://www.vics.or.jp/english/
[2] Masatoshi Arikawa, Takeshi Sagara, Koji Okamura, “Spatial Media Fusion Project,” Proceedings of 2000 Kyoto International Conference, on Digital Libraries: Research and Practice, pp.75-82, 2000.
[3] Shobunsha Publications Inc.,<http://www.mapple.co.jp/>.
[4] EZnaviwalk, KDDI au, <http://www.au.kddi.com/>.
[5] Japanese Morphological Analysis System *ChaSen*, <http://chasen.aist-nara.ac.jp/hiki/ChaSen/>.

The concept of relevance in mobile maps

Tumasch Reichenbacher

Department of Cartography, Technical University of Munich
tumasch@bv.tum.de

Abstract

Mobile map based services and location-based services (LBS) are going places. Yet, they face acceptance problems for such reasons as lacking user focus and relevance. This paper looks at the concept of relevance and its potential for and application to mobile cartography. First, theoretical concepts of relevance from other disciplines are studied followed by an analysis of different approaches for the assessment of relevance. A basic model for the determination of geographic relevance is postulated and the application modes of relevance to mobile cartography are elaborated. Two applications of relevance to mobile cartography are discussed with an emphasis on visualisation techniques for the presentation of relevance in maps. Some final remarks conclude the comments on relevance in mobile maps.

1 Introduction

Mobile usage of digital geographic information is becoming mainstream technology. Several types of applications and services have evolved over the last few years and are already broadly used, e.g., car navigation systems, tour guides, and, of course, location-based services (LBS). However, research is still sparse when it comes to developing appropriate visualisations of geographic information for small displays of mobile devices. The major challenge for mobile cartography is the separation of relevant and irrelevant information by finding an acceptable degree of information reduction to the relevant, i.e., presenting as much information as needed and as little as required.

Relevance is becoming an important feature for mobile services in general and especially for LBS and map based mobile services. Raper et al. (2002), for example, claim that "… understanding the individual 'geographical relevance' of information will be necessary for location-based services to provide appropriate information …". Geake (2000) mentions among other criteria for successful LBS the *relevance* of delivered information. Similarly, Oinas-Kukkonen and Kurkela (2003) propose seven key design principles for highly goal-driven mobile services: (1) provide information addressing the needs of users on the move (mobility), (2) make life easier (usefulness), (3) include relevant information (relevance), (4) simple and easy to use (ease of use), (5) most important information should be the easiest to locate (fluency of information), (6) focus on user's terminology and navigational structure (user-centredness), and (7) they should be adapted to each and every user's own needs and capabilities (personalization) or in short: "A good mobile service provides additional value for the user and is fast and natural to use".

This paper describes the more fundamental concept of relevance and its application as well as significance to cartography in general and mobile cartography in specific. The major goal of this research is the adequate adaptation of geographic information and its presentation in order to reach a higher degree of relevance for the user as outlined in (Reichenbacher 2005) and to stress the significance of relevance for a successful implementation of mobile cartography in commercial products and services.

The remainder of the paper will examine the concept of information relevance and its role in visualisation. First, theoretical concepts of relevance from other disciplines are examined, followed by a discussion of different dimensions of relevance and their relation to the usage context. In a further step, approaches for the assessment of relevance are analysed serving as a base for a simple method for the calculation of compound geographic relevance values. Moreover, the possible applications of relevance values to mobile cartography are highlighted. Finally, the visualisation of relevance in mobile maps and possible implementations are demonstrated on behalf of map examples.

2 The concept of relevance in other disciplines

Relevance stems from the Latin word *relevare* meaning to raise up, to relieve. Although relevance is a rather fuzzy concept, as humans we intuitively know what relevance is even without having a clear concept of it. In general usage we apply it synonymously for importance or pertinence. Saracevic (1996) offers a universal and useful definition of relevance derived from its general qualities: “ […] relevance involves an interactive, dynamic establishment of a relation by inference, with intentions toward a context. […] relevance may be defined as a criterion reflecting the effectiveness of exchange of information between people (or between people and objects potentially conveying information) in communicative relation, all within a context.” Depending on the relations established Saracevic (1996) distinguishes five manifestations of relevance (see also Fig.2): (1) the system or algorithmic, (2) the topical, (3) the cognitive, (4) the situational, and (5) the motivational relevance that form together a system of relevancies on different levels. However, the concept of relevance has some distinct notions in different scientific disciplines. A major distinction has to be made between *objective* (1) and *subjective* (2-5) relevance.

The former has been the one used for years in information retrieval (IR) where the goal is to “… retrieve all the relevant documents [and] at the same retrieving as few of the non-relevant documents as possible” (van Rijsbergen 1979, p. 6). The efficiency and effectiveness of this retrieval process is generally determined by the measures *precision* and *recall* (Mizzaro 2001). These measures reflect the notion of binary relevance, i.e. the document is either relevant or not. The assumption is that the success, i.e. the relevance of the documents, can be determined independently from the user by a system, hence the relevance is *objective*. However in practice, as it has been pointed out by many researchers (c.f. Saracevic 1996; Wilson and Sperber 2004; Cosijn and Ingwersen 2000), there are grades of relevance rather than a binary relevance and hence degrees of relevance need to be introduced. This is usually done by ranking the retrieved documents based on the similarity of the document and the query. All Internet search engines (e.g., Google) have an implemented ranking mechanism based on the cosine measure (the scalar product of each document vector against the query vector defining the presentation order of the retrieved documents).

In linguistics, pragmatics, and communication science the objective relevance is not sufficient or useful. These disciplines focus on *subjective* relevance. The basic assumption is that the relevance of entities is largely determined by the user. An important contribution is the relevance theory of communication proposed by (Sperber and Wilson 1986; Sperber and Wilson 1995; Wilson and Sperber 2004). The theory is based on cognitive psychology understanding relevance as a judgement criterion during cognitive processes, the criteria for inputs to be relevant to an individual being *effect* and *effort*. The underlying assumption is that human cognition is directed towards maximising relevance. Sperber and Wilson (1995) define relevance as follows: "An assumption is relevant in a context if and only if it has some contextual effect in that context […] An assumption is relevant in a context to the extent that its contextual effects in this context are large […] An assumption is relevant in a context to the extent that the effort to process it in this context is small." Some of these findings can help to grasp the concept of relevance for geographic information, although the extension to the spatial domain presents an even higher degree of complexity.

3 Relevance in mobile cartography

3.1 Relevance and Context

The discussion of relevance in section two reveals that the concept of relevance is very much dependent on context, since the context shapes the boundary between relevancy and irrelevancy. Relevance always exists in relation to a specific context and a change of context alters the quality of being relevant. Recent research work in the domain of LBS, mobile cartography, and mobile web services is rich on investigations about context-awareness; see for example (Reichenbacher 2004; Nivala and Sarjakoski 2003; Dey and Abowd 1999; Schmidt et al. 1998). As pointed out by the author on other occasions more than the mere spatial relevance is required for suitable adaptation of mobile maps as well as a success of the relevance principle in mobile cartography. Apart from the spatial relation there are other factors and challenges in mobile usage situations originating from physical environmental states, temporal constraints, mobile users' information needs and activities, technical limitations and many more. These factors necessitate an optimal exploitation of the limited map space on mobile devices, i.e. a stringent selection of relevant objects to be displayed. The presented geographic information must be relevant to the context of use, i.e., a mobile map should satisfy the user's contextual information needs and adequately present the respective information objects.

3.2 Relevance types

By combining the manifestations of relevance mentioned above and the contextual factors of importance in mobile environments a set of relevance types for mobile cartographic applications and services can be derived (Fig. 1). The depicted relevance types describe the different relations between geographic information objects and context dimensions. Accordingly objective relevance types that are independent from the user and subjective relevance types related to personal aspects of the user are separated.

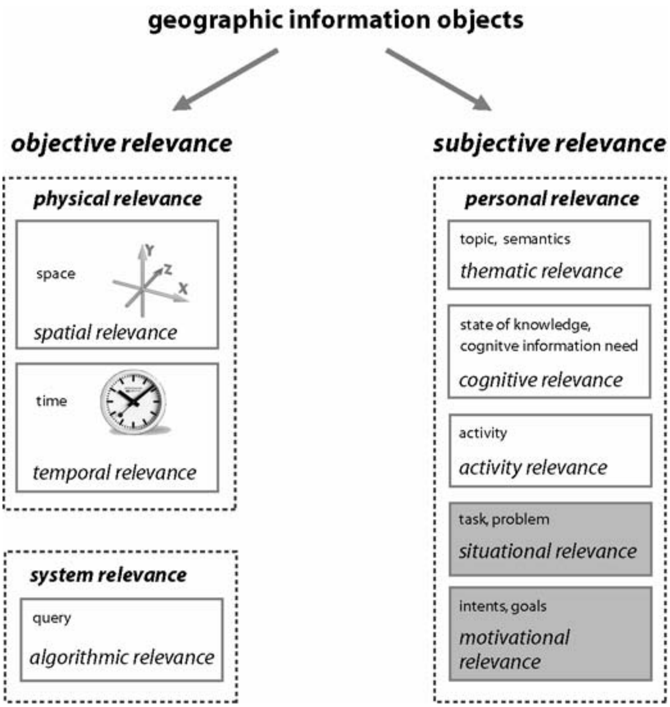


Fig. 1. Relevance types for mobile cartographic applications (based on Saracevic, 1996)

An extension of the manifestations of relevance proposed by Saracevic (1996) is the introduction of activity relevance as a super-class of situational and motivational relevance (shaded in grey in Fig. 1). These relevance dimensions represent two different aspects of an activity: the motivation and the embedment in a situation. Yet, it will not always be possible to clearly separate all these relevance types depicted in Fig. 1, since in practice they often might overlap to some extent.

4 Relevance assessment and relevance measures for mobile cartography

The objective of relevance assessment is to formalise the relation between geospatial objects and the different dimensions shown in Fig. 1. Intuitively we might propose some general rules of thumbs for the assessment of a hierarchy of relevance for geospatial objects:

- nearer objects are more relevant than objects further away
- visible objects are more relevant than hidden or invisible objects
- audible objects are more relevant than inaudible objects
- an object or an object's attribute is relevant, if it is needed for the successful completion of an activity
- objects that are linkable to users' prevalent knowledge are more relevant
- objects with low information content for the user are less relevant

Some of these rules can be formalised as will be discussed below. However, most formal approaches for relevance assessment deal mainly with non-spatial information. Nevertheless, the assessment methods can to some extent be adapted for geospatial information.

4.1 Utility functions

Utility is an economical measure for satisfaction gained from products or services. Utility functions are a widely used tool in economics to model user preferences. A utility function $u : X \rightarrow R$ maps attributes of objects to preferences by assigning scores to alternatives in the set X (http://en.wikipedia.org/wiki/Utility; accessed 13/11/2005). If x and y are alternatives then $u(x) > u(y)$ states that x is preferred to y. An example of a simple function to model the utility of features dependent on their distance could look like the one depicted in Fig. 2.

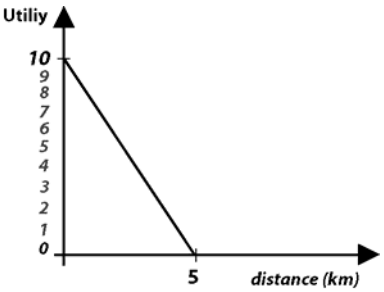


Fig. 2. A simple utility function for spatial distance of objects

This function can look differently in different contexts, i.e. it can be adapted for diverse activities. In addition it can be applied to other relevance dimensions as well, i.e. other kinds of distance values are used.

4.2 Information retrieval functions

Classic IR applies a vector space model for ranking documents according to their *similarity* to a query. IR methods relate the content of documents to the query that represents the user's information need. Classic document retrieval uses methods of semantic distance to evaluate the match of topic. Such techniques can also be used for the selection of topically relevant geospatial features based on their attributes. For a method of defining thematic distances see for example (Jones et al. 2001). An alternative approach is probabilistic IR aiming at ranking documents according to the evaluated *probability* of relevance to the user's information needs (Crestani et al. 1998).

4.3 Fuzzy sets

Fuzzy set theory proposes a way to model elements that do not belong to a set in a binary way, i.e. is or is not a member of the set. Instead different membership functions can be applied to define the degree of membership to a specific set. A membership function for a fuzzy set A looks like $\mu_{\mathrm{A}}(x) \in [0,1]$, i.e. x is assigned a value between 0 and 1. Typical membership functions are the triangle or the trapezoid function. Schmidt and Gellersen (2001) propose a model of context validity based on fuzzy set theory that can be easily altered for purposes of geographic relevance. The approach models the spatial and temporal decrease of validity of context in relation to its origin. In the same manner spatial and temporal relevancy can be modelled fuzzily, i.e. in a non binary way.

4.4 Observation-based approaches

The relevance of geospatial objects can also be derived from observing user behaviour. From the knowledge about what feature types are typically associated with and hence relevant for specific activities an enumerative list of typical activity-feature relations can be elaborated. The spatial range of the activity further defines the set of the spatially relevant objects for that activity. An alternative and more advanced approach is the logging of movement traces and feature selections that can serve as examples for learning algorithms to induce rules for feature selection or as the base for calculating a relevance score based on the observed frequency of requested features for a specific activity pattern. Attempts of modelling spatio-temporal behaviour of users for LBS are described in (Mountain and Raper 2002).

4.5 Geographic information relevance assessment

Approaches to formalise relevance or importance of geospatial information are rather rare (Zipf 2003; Reichenbacher 2004; Reichenbacher 2005). The calculation of isolated relevance for the corresponding relevance types might improve the utility of services, but generally interdependencies exist between the single relevance dimensions and a concurrent determination of a compound relevance factor is certainly advantageous. An example of calculating such a compound relevance factor for events based on this approach is described in (Reichenbacher 2004). Individually calculated spatial, temporal, and thematic relevance values are summed up to a total relevance factor indicating the relevance of an event to the user with a specific topic interest at a certain location and time.

The spatial relevance is modelled as a function of spatial distance of the objects (O) to the current location (L): $R(O) = f\left(dist\,\overline{OL}\right)$. Likewise temporal relevance can be modelled as a function of the temporal distance between the time of usage and a time reference of an object or event: $R(O) = f\left(dist\,\overline{t_O t_L}\right)$. Thematic relevance can be modelled as a function of the distance between feature attribute values and query terms organised as concepts (C) in an ontology. Then a semantic distance function could look like: $R(O) = f\left(dist\,\overline{C_O C_q}\right)$. Such methods for computing semantic similarity measures are described for instance in the SPIRIT project (Jones et al. 2001). Simplistically the general relevance R for an object O_i can be modelled as a function: $R(O_i) = \sum_{j=0}^{n} w_j \bullet r_j$; where j is the relevance type (spatial, temporal, …), r_j is the value for the relevance type j, and w_j is the weight for the value of j.

The setting of weights for the considered relevance types is dependent on context. In practice not all relevance types have to or can be considered in the function. Yet, the consideration of more than one relevance type produces a more differentiated picture of object relevancies (see Reichenbacher 2005).

Of course, the proposed calculation is oversimplified and does so far not take into account mutual influences of the single relevance dimensions. In addition, apart from the spatial distance to the user's position spatial relevance might become apparent as (1) an area of interest (AOI) and (2) an activity or social space. Correspondingly temporal relevance (timeliness) can express itself as (1) a time period or interval (e.g. duration, day, or month) in which a point of interest is relevant or an activity is practicable and (2) as reachability (means of transport/timetable; speed; constraints) or (3) as qualitatively or socially defined times (e.g. leisure time).

5 Applications of relevance in mobile cartography

The application of relevance to mobile cartography is twofold: 1) relevance of geospatial features in relation to a specific usage situation may be used for selecting or filtering geospatial information from geodatabases. 2) relevance measure values can be applied to the map graphics to visually encode the differences in relevance of objects. Both applications should improve the usability substantially by increasing the relevance of the presented information. It has to be stressed here that the two different applications reflect together the significant distinction between the terms *relevance* and *saliency*. The former relates to geospatial objects in the (geospatial) *information space*. The latter refers to map objects in the *visual/symbolic map space*. The main task of cartography is to select as many relevant objects as possible and portray them adequately as salient features in a map (Dent 1999). This construction of relevance for mobile usage situations is achieved through personalisation which is not equally feasible with static map products (Fig. 3).

Classic static maps have to supply a greater quantity and more general kind of information to meet different information needs not known a priori. Mobile, personalised maps on the other hand can be based on queries that express an information need and can be much more tailored to specific usage contexts. By that they tend to include more of the relevant and fewer of the irrelevant objects.

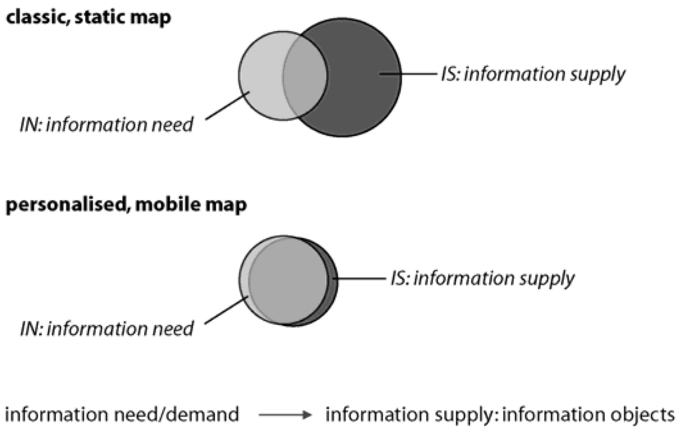


Fig. 3. Relevance enhancement through map personalisation

The design methodology for mobile maps follows quite tightly the thematic mapping approach. The two main parts of the map are the base map layer and the thematic layer. General and universal information that is more or less equally relevant for all kind of mobile maps is collected in the base map layer. This layer holds the information necessary for orientation and spatial reference. However, referential information can also be selected in a hierarchical manner as distinct levels of detail.

On top of the base map layer specific features relevant for the usage context are composed to one or several thematic layers. The content of such a layer is dependent on the spatial problem to be supported by it. For localisation support the user's positions in the map and visible landmarks have to be shown. For way finding support possible routes and route specific landmarks need to be included in the map. In addition features matching the user's interests (e.g. historical sites), special needs (e.g. WC for handicapped) or role (e.g. tourist) might be contained along the route. To support user activities the relevant features for a specific activity are portrayed in the layer. Features that are of general relevance to the activity can be separated from additional features that are user specific for that activity. For instance, for the activity *biking* the thematic layer might show bike routes depending on user specific characteristics or interests, e.g. young – old, fit – unfit. Features matching the general topic of interest to the user form thematic map layers. The thematic information is either presented as individual points of interest (POI) or areas of interests (AOI) that might also be clusters of POI.

The producing of mobile maps starts with filtering geographic information based on relevance. In a first step spatial filtering is accomplished by pre-selecting those features included in an AOI, for instance in an activity zone. The application of buffers is an alternative method for pre-selecting relevant objects for specific locations or along routes.

The parameters for a geographical object query can be derived from the context parameter values. The OGC Filter Encoding Implementation Specification (OGC 2001) offers a wide range of operators to build filter expressions for queries. Spatial operators are used for comparing the spatial relations between the query and the features (e.g. *within*, *overlaps*, *intersects*, *contains*, *DWithin*, *BBOX*, etc.). The non-spatial properties can be filtered applying comparison operators. Both types of operators can be composed to more complex filters using logical operators. Temporal relevance filtering can be based on feature information, if a temporal relation is stored as an attribute of the feature. Thematically relevant objects can be filtered by comparing feature attributes with query terms.

Alternatively features might be filtered based on a pre-calculated normalised compound relevance measure values that are stored as an attribute. The selection of relevant objects against non-relevant would then be determined by the setting of a relevance threshold value, e.g. retrieve all objects with $R > 0.75$.

As explained above there is a difference in relevancy and appropriateness or adequacy of the presentation. A high value of information relevance, i.e. feature filtering, alone does not necessarily lead to a high degree of relevance of visualisation where cartographic knowledge is essential (Fig. 4).

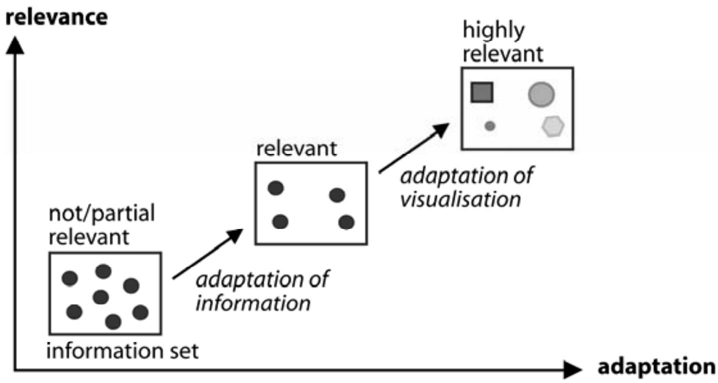


Fig. 4. Relevance enhancement through adaptations

The overall relevance of a presentation of geographic information might not fully exploit its potential. Through different adaptations in the visualisation domain a higher total relevance of a mobile map service can be realised (Reichenbacher 2004). The next chapter illustrates possible techniques of visualising differences in relevance in mobile maps.

6 Relevance visualisations

Mobile map based services aim at presenting geographic information with the highest possible relevance to the user. First of all the selected or filtered objects need to be appropriately symbolised. An augmentation to the service utility is the visual presentation of the relevance of the objects that can be more efficiently processed when synoptically perceived.

One way to visualise relevance in a mobile map is the direct mapping of relevance values to graphical variables, i.e. the attempt to transform relevance values to salience cues. Thereby either binary relevance, i.e. relevant and non-relevant, or degrees of relevance (relevance order, grades of relevance) can be encoded. The cartographic toolbox is rich in techniques for putting visual emphasis or focus on important features or for emphasising their relevance respectively. Some examples are listed below and also illustrated in Figure 5:

- highlighting the object using a different (maybe more luscious) colour (*colour*, *hue*)
- emphasising the object using brighter colours (*value*)
- focussing the object while blurring the other objects (*clarity, focus, crispness*)
- decreasing the opacity of the object against the other objects (*opacity*, *transparency*)
- emphasising the symbol size or the size of the outline (*size*)

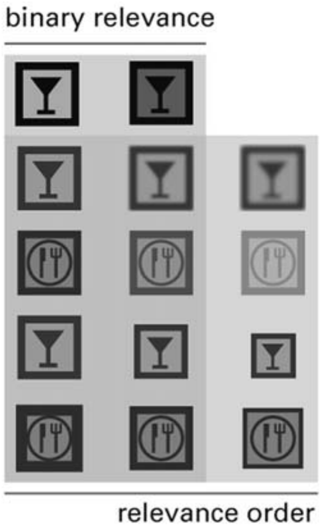


Fig. 5. Examples of feature emphasising techniques usable for relevance visualisation

Moreover dynamic variables such as *duration*, *frequency*, or *order* can be applied to animations of the classic graphical variables to attract the user’s attention, i.e. to show for instance the most relevant object among relevant objects. For these kind of examples like colour transformations, symbol rotation or blinking symbols, or increasing and shrinking, as well as further examples of application of different graphical variables and metaphors (hotspot metaphor, fade-away metaphor) to the relevance visualisation of POI and events see (Reichenbacher 2005). The relevance of POI can intuitively be visualised as different grades of opacity values for the POI symbol (Fig. 6 left). Generally the features of the base map are symbolised in grey or in very light colours as shown in Fig. 6 left and right. However, if single feature classes or features are highly relevant for an activity or context they can be emphasised by showing them in colours commonly used. In the case of an intended ‘walk through the nature parts of the city’ the green areas and water courses and bodies are relevant and would be specially accentuated (Fig. 6 middle). A possible way to intuitively convey the relevance of different AOI is the cold-hot metaphor. Different grades of blue and red hues are picked to present the hotness or coolness of an area (Fig. 6 right).

Fig. 6. Relevance of POI encoded with opacity (left); colour-grey contrast (middle); hot-cold metaphor (right). (© base data: *Städtisches Vermessungsamt München*)

Although the usability of mobile map based services is likely to be improved by incorporating relevance, it poses at the same time new problems: so far it is not clear which graphical variables are more apt for efficiently visualising relevance. In addition, the meaning of relevance symbolisation encoded is not always obvious for the user. Colours of map symbols and relevance symbolisation might disturb each other or colours might be already taken by map symbols and may no longer be available for relevance visualisation.

Technically relevance filtering and visualisation can be implemented within an adaptive geovisualisation service as described by (Reichenbacher 2004). The service sends requests to a *Web Feature Server* (WFS) which in turn delivers features encoded in GML that can be transformed to *Scalable Vector Graphics* (SVG) encoded maps through *Extensible Stylesheet Language Transformation* (XSLT). This approach offers various possibilities for bringing more relevance into the service. First, filter operations can be included in the WFS request and second the XSL transformation allows for the mapping of context derived relevance values to graphical variables which are mostly easily codable as SVG elements.

7 Conclusions

This contribution aimed at unveiling the benefits of a deeper understanding of the relevance concept for mobile cartography and illustrated that applying relevance values to mobile geovisualisation services can enhance the service usability. Further research should focus on connecting typical activities of mobile users, their embedment in contexts with fitting feature sets and the most suitable presentation forms for these activities. The general feasibility and performance efficiency of relevance computation must be proved before an application to mobile services can make sense, since time is one crucial factor in mobile services. Further enhancements should focus on interdependencies of the single relevance dimensions and a more differentiated formalisation of the relevance assessment. The effects of relevance filtering and visualisation within map based mobile services need to be tested in user studies with everyday tasks.

References

Cosijn E and Ingwersen P (2000) Dimensions of relevance. Information Processing and Management, 36:533-550

Crestani F, Lalmas M, van Rijsbergen C J and Campbell I (1998) "Is This Document Relevant? ... Probably": A Survey of Probabilisitic Models in Information Retrieval. ACM Computing Surveys, 30(4): 528-552

Dent B D (1999) Cartography. Thematic Map Design. Wm C Brown, Dubuque

Dey A K and Abowd G D (1999) Towards a Better Understanding of Context and Context- Awareness. Technical Report. GIT-GVU-99-22. Georgia Institute of Technology. Atlanta (GA)

Geake V (2000) Web mapping. Get mobile, it's the place to be. GIS Europe(09/2000) http://www.geoplace.com/ge/2000/0900/0900mob.asp

Jones C B, Alani H and Tudhope D (2001) Geographical Information Retrieval with Ontologies of Place. In: D R Montello (ed) Spatial Information Theory: Foundations of Geographic Information Science. International Conference, COSIT 2001. Morro Bay, CA, USA, September 19-23, 2001, Proceedings. Lecture Notes in Computer Science 2205. Springer, New York, pp 322-335

Mizzaro S (2001) A New Measure Of Retrieval Effectiveness (Or: What's Wrong With Precision And Recall). In: T Ojala (ed) International Workshop on Information Retrieval (IR'2001). Infotech Oulu, Oulu, Finland, September 19-21, 2001, pp 43-52

Mountain D and Raper J (2002) Modelling human spatio-temporal behaviour: A challenge for location-based services. Proceedings AGI 2002

Nivala AM and Sarjakoski LT (2003) Need for Context-Aware Topographic Maps in Mobile Devices. Proceedings 9th Scandinavian Research Conference on Geographic Information Science ScanGIS’2003, Espoo, Finland

OGC (2001) Filter Encoding Implementation Specification. Open GIS Implementation Specification, Project Document OGC 02-059, Open GIS Consortium

Oinas-Kukkonen H and Kurkela V (2003) Developing Successful Mobile Applications. Proceedings International Conference on Computer Science and Technology (IASTED), Cancun, Mexico, May 19-21, 2003, pp 50-54

Raper J, Dykes J, Wood J, Mountain D, Krause A and Rhind D (2002) A framework for evaluating geographical information. Journal of Information Science 28(1): 39-50

Reichenbacher T (2004) Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices. Verlag Dr Hut, München

Reichenbacher T (2005) The Importance of Being Relevant. Proceedings XXII International Cartographic Conference, A Coruna, Spain, July 9-16, 2005

Saracevic T (1996) Relevance reconsidered. Proceedings Second Conference on Conceptions of Library and Information Science (CoLIS 2), Copenhagen, October 14-17, 1996, pp 201-218

Schmidt A, Beigl M and Gellersen HW (1998) There is more to context than location. Proceedings International Workshop on Interactive Applications of Mobile Computing (IMC98), Rostock, November, 1998

Schmidt A and Gellersen HW (2001) Modell, Architektur und Plattform für Informationssysteme mit Kontextbezug. Informatik Forschung und Entwicklung(16): 213-224

Sperber D and Wilson D (1986) Relevance: Communication and cognition. Harvard University Press, Cambridge (MA)

Sperber D and Wilson D (1995) Relevance: Communication and cognition. Blackwell, Oxford

van Rijsbergen C J (1979) Information Retrieval. Butterworths, London

Wilson D and Sperber D (2004) Relevance Theory. In: L Horn (ed) Handbook of Pragmatics. Blackwell, Oxford, pp 607-632

Zipf A (2003) Die Relevanz von Geoobjekten in Fokuskarten - Zur Bestimmung von Funktionen für die personen- und kontextsensitive Bewertung der Bedeutung von Geoobjekten für Fokuskarten. Proceedings Symposium für Angewandte Geographische Informationstechnologie AGIT 2003, Salzburg, Austria

A Knowledge-Based Map Adaptation Approach for Mobile Map Services

L. Tiina Sarjakoski, Tommi Koivula, Tapani Sarjakoski

Finnish Geodetic Institute
P.O. Box 15, FIN-02431 Masala, Finland
tiina.sarjakoski@fgi.fi, tommi.koivula@fgi.fi, tapani.sarjakoski@fgi.fi

Abstract

In order to be able to utilise and manage various map specifications for mobile maps, a Map Specification Knowledge Base approach is introduced and formally described. A Map Specification Knowledge Base is a set of specification rules, properties and assignments constituting the system's knowledge about the map to be delivered to the mobile user. Such knowledge includes rules connecting context parameters with the parameters for the level of details of the map and map generalisation, and settings for the map visualisation. The method uses an object-oriented approach to carry out the multiple-inheritance of the map specification modifiers based on the importance of the rules. The implemented map service can deliver various types of maps, which match the current context parameters and user preferences in real-time. To provide third-party mobile service developers with the possibility to easily create and modify map specifications for their specific needs, a web-based editor, called Map Specification Tool was implemented. The research described here was part of an EU- project referred to as GiMoDig (Geospatial info-mobility service by real-time data-integration and generalisation). A prototype of a cartographic map service was implemented to deliver geospatial data to mobile users in real-time, and is based on emerging standards, such as

Extensible Mark-up Language (XML) and Open Geospatial Consortium's (OGC) specifications.

1 Introduction

1.1 Background

The evolving map applications for different types of mobile devices present new challenges for mobile services and application providers. Out of all the mobile applications, the applications that use the Global Positioning System (GPS) and display the map on a mobile device, are the most in demand at the moment. Thus, the national topographic databases play a key role in applications that need detailed geospatial information. On top of the topographic data provided by the geospatial data services third-party service developers may create various end-user applications and display so called Points of Interest (PoI) data. At the same time, the developers of mobile applications face new challenges when aiming to meet the demands of the various user groups of mobile services. One of the crucial challenges faced is as how a single map service could be used to deliver and adapt mobile maps for different kinds of users in different usage situations.

The research described here is based on the results from an EU-funded project carried out in 2001-2004 called GiMoDig (Geospatial info-mobility service by real-time data-integration and generalisation). In the project, a prototype of a cartographic map service was implemented to deliver geospatial data for mobile users in real-time. As the overall objective of the GiMoDig project was to improve the accessibility and interoperability of national topographic databases, the data was delivered from the geo-databases in the participating countries' National Mapping Agencies (NMAs) (Finland, Sweden, Denmark and Germany), and resulted in a vector-formatted high quality Scalable Vector Graphics (SVG) map being displayed on the user's mobile device. The GiMoDig service was based on emerging standards, such as Extensible Mark-up Language (XML) and Open Geospatial Consortium's (OGC) specifications.

The focus of this research is to describe a mobile map service approach that includes a knowledge base which is used when the service controls in real-time the map adaptation of mobile maps. The chapter is based on an earlier paper by the authors (Sarjakoski et al., 2005).

1.2 Previous research

Today, the ability for adaptation and context sensitiveness are regarded as essential characteristics of Location-Based Services (LBS) (Meng et al., 2005). According to Oppermann, (1994) an adaptive system is capable of changing its own characteristics automatically according to the user needs. For example, Baus et al. (2001) describe a system that takes the location of the user and adapts the presentation of route directions according to the characteristics of the user's mobile device as well as to the cognitive resources of the user. Dey (2001) defined a system to be a context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. Reichenbacher (2001, 2004) was among those who started to study the process of the adaptive and dynamic generation of map visualisations for mobile users in relation to cartographic principles. By using the context information available, mobile map services could adapt the map visualisation and functionality for different usage situations and individual user needs (Sarjakoski and Nivala, 2005).

Cartographic expert systems were intensively studied at the beginning of 90s, for tasks such as cartographic design (Müller and Zeshen, 1990; Forrest, 1993), or map generalisation (Buttenfield and McMaster, 1991). However, the low success rate of such systems was claimed to be due to the lack of the necessary rules or their incompleteness. There has recently been an increasing interest in the research on cartographic expert systems (e.g., Stefanakis and Tsoulos, 2005). Techniques from areas such as the object-oriented approach, constrained or agent-based technology are combined with technology from previous expert systems. In our approach we have adopted some object-oriented techniques to the knowledge-based approach, while attempting to find a solution as how to use only a single map service to deliver different kind of maps and adapt them to various user needs (Sarjakoski and Sarjakoski, 2005).

The following section briefly describes the GiMoDig service architecture, followed by a formal description and examples of a Map Specification Knowledge Base (MSKB). The implementation of the Map Specification Tool Editor is described in Section 3. Finally, the conclusions are given.

2 The GiMoDig map service

The GiMoDig service prototype incorporates a number of geo-spatial tasks: data transformation to a common reference frame, thematic harmonisation of geodata according to a global schema, real-time generalisation and the integration of Points of Interest (PoI) data, and applications tailored to meet a range of needs for different mobile users.

The GiMoDig prototype is based on layer-based architecture, consisting of six layers (GiMoDig, 2005; Lehto and Sarjakoski, 2005). On the first level, the data providers (i.e. in GiMoDig, the NMAs) run a Data Service providing raw spatial data in an XML-encoded form, Figure 1. Above the data services is the Data Integration Service layer, which takes care of e.g. coordinate transformations to a common EUREF- coordinate system, and schema transformations. On the third level in the architecture is the Data Processing layer for processing tasks such as map generalisation or dynamic map labeling. The fourth layer in the system architecture is called Portal Service. The main responsibilities of this layer include processing the service requests coming from the client, subsequently forwarding the request in an appropriate form to the Data Processing layer below, and transforming the resulting piece of geospatial data into a visual representation, according to the capabilities of the client platform in question. It should be noted that in the service architecture the query results are represented in the form of XML-encoded spatial data (i.e. Geography Markup Language (GML) up to the Portal Layer. In this layer the query dataset is transformed into a visual map image (i.e. an SVG map) and styled appropriately for the client environment in use.

The fifth layer is the so-called Value-Added Service (VAS) layer. The main task of this layer is to control the creation process of the map that will be delivered to client applications, taking into account the parameters related to the adaptive map display. It determines, in particular, the content of the PoI data that is overlaid on the topographic data, and also the selection and style of the topographic data. The service access on the VAS layer is based on a proprietary, use case-specific query interface.

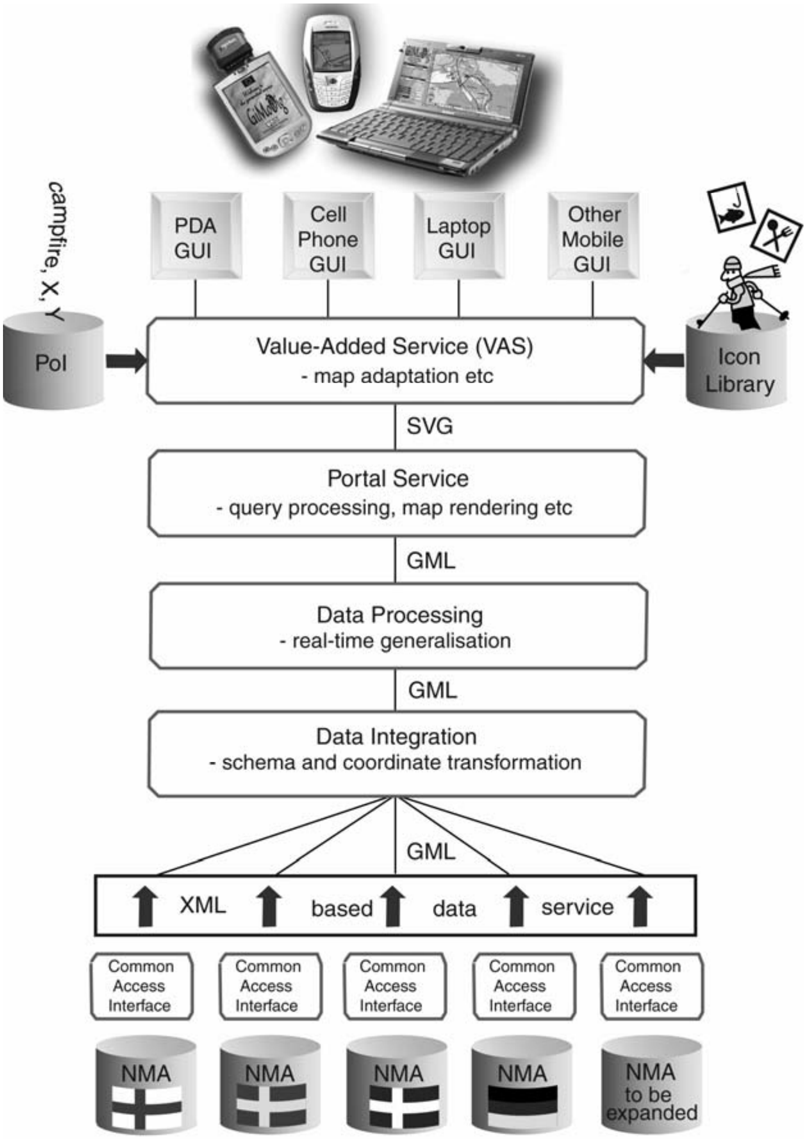


Fig. 1. The GiMoDig layer-based service architecture (modified from Sarjakoski and Sarjakoski, 2005).

One of the main research topics in this respect is the use of context parameters for adaptive maps for different kinds of mobile users. These parameters include the following detailed information being used in the query:

- *usecase* to indicate the activity the requestor is currently involved in
- *lod* (Level of Detail) to define the resolution of the map representation
- *time* for indicating the temporal context of the usage situation
- *age* group of the user to affect the map visualisation process based on the user's age group
- *device* to set the context parameters related to the physical device currently in use
- *center* to indicate the location of the map to be requested
- *scale* to set the display scale
- *position* to set the user's current position

Examples of the value sets of the parameters in a query are listed in Table 1.

Table 1. Context parameters and their value sets selected for the GiMoDig prototype.

Parameters	Value sets v_{pi}
p_{USE_CASE}	{"outdoors", "cycling", "emergency", "expert", o}
p_{LOD}	{"detailed", "basic", "intermediate", "overview", "egm", o}
p_{TIME}	{"winter", "spring", "summer", "autumn", "day", "night", o}
p_{AGE}	{ 0-10, 11-17, 18-45, 46-, o }
$p_{LANGUAGE}$	{"English", "Finnish", "Danish", "German", o}
p_{DEVICE}	{"ipaq", "pc", o}

The Value-Added Service contains a knowledge base for different contexts and visualisations assigned to them. The VAS also uses an SVG-based icon library and a Points of Interest (PoI) database, which were developed in the GiMoDig-project. The icon library consists of icon sets in four different styles (totally 161 icons), which are selected depending on the age group of the user for the requested map (Sarjakoski et al., 2003; Nivala and Sarjakoski, 2005). The PoI-database contains coordinates of the PoIs (Figure 1).

The client applications consisting of four use-case-based applications are on the sixth layer: A Hiker in a National Park (so called Outdoors use case), Cycling and Emergency use cases, and a tool for expert users to tailor mobile maps (Sarjakoski and Sarjakoski, 2004, 2005). An advantage of the layer-based architecture approach is that the results can be adapted to various client environments. Three client platforms were considered for implementation: traditional web browsing on a portable PC platform, more restricted web access on Personal Digital Assistant (PDA) devices and a client application on mobile cell phones.

3 Map Specification Knowledge Base (MSKB)

One of the research problems faced in the GiMoDig project was how a single map service could be used to deliver and adapt mobile maps for various users in different usage situations. In order to be able to utilise and manage various map specifications for the different maps, a Map Specification Knowledge Base approach was needed. A Map Specification Knowledge Base is a set of specification rules, properties and assignments constituting the system's knowledge about the map to be delivered to the mobile user. Such knowledge includes user preferences for the mobile context, parameters for the level of details on the map and map generalisation, and settings for the map visualisation (styles, colours) and map contents (i.e., displayed map feature classes and icons), etc. However, the large number of combinations of the context-parameter values, which is several thousands already in our limited example (Table 1), would make it nearly impossible to create specifications one by one. To solve this problem and to minimise the number of specification rules and properties that needs to be entered manually, a method that uses multiple-inheritance of modifiers based on the importance of the conditions is introduced in the following paragraphs. The most common properties (such as "styles of the map") are mapped for the most important conditions (such as "a use case") and then inherited to other less important conditions. The innovative aspect of the Map Specification Knowledge Base and the map specification approach is that the service can deliver various types of maps, which match the current context parameters and user preferences in real-time. In the following paragraphs we give a formal definition of the map specification knowledge base, and we also present an algorithm that is used in run time for finding map specifications matching with the current context parameters.

3.1 Context and context parameters

In general, context can be seen as a collection of information to characterise the situation of an entity. Entity is a person, place or an object that is relevant to the interaction between a user and an application, including the user and the applications themselves (Dey and Abowd, 1999). Furthermore, Dey (2001) defined a system to be a context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. Nivala and Sarjakoski (2003) identified in their study mobile contexts based on user tests of topographic maps in mobile devices. These included location, system, purpose of use, time, physical surroundings, navigational history, orientation, user and cultural and social elements. They proposed that embedding context awareness into the topographic maps in mobile devices could increase the usability of the map service better to support its user.

In location-based services, context can also be seen as a mean to integrate physical world and the information space together (Gross and Specht, 2001). In this way the problem can be formalised into a question how the entities from the physical world should affect the properties of a map.

To be able to utilise the context, first we formalise it into a set of context parameters, which identifies the context itself. Context can be formalised by using dimensions (Tomai and Kavouras, 2005). Dimensions are specific variables when combined to give meaning to information. They describe the surroundings for all the user actions as well as the user application used. Using n dimensions, we define context C as

$$C=(p_1,\ldots,p_n) \quad \{ v_{p1},\ldots,v_{pn} \mid p_i \in v_{pi} \}, \qquad (1)$$

where p_i is the context parameter and v_{pi} is a set of possible values of context parameter p_i. All sets v_{pi} also contain a value that indicates that the actual value of this parameter is unknown. We denote this value as o.

The characteristics of a context can be either static or dynamic (Goulimis et al., 2003). Parameters like device, user, cultural and social elements can be considered to be static, since they probably will not change during the user activities. Parameters like time, history and location are considered dynamic. Goulimis et. al. (2003) categorised the associations for entities into: sensed associations, derived associations and profiles associations. Similarly, in terms of our definitions dynamic context parameters can be:

- *Sensed parameters* that are read from the sensors of the mobile device. For example, location could be sensed from the GPS-module.
- *Derived parameters* that are derived from the other information. For example, a possibility to ski could be calculated from time, location and local snow statistics.
- *Adapted parameters* that profile complex characteristics using simple variables. For example, information about the user can be given by the parameters like age, gender and nationality.

The following example illustrates how contexts are used. Let say we have a context $C=\{p_1,\ldots,p_5\}$ and the values of the context parameters can be the following: v_{p1}={winter,summer,o}, v_{p2}={pc,handheld,phone,o}, v_{p3}={male,female,o}, v_{p4}={age as integer }∪{o} and v_{p5}={lod_basic,lod_overview,o}. One possible realisation of context C could be C' = (o,handheld,o,25,o). It simply states that the user has a handheld device, his/her age is 25, and the gender and season are currently unknown.

In the algorithm to map a context into a set of map properties (which are introduced in Section 3.3) the order of the parameters in a context C is also important. The context is evaluated so that the first parameters are considered to be more important than the last ones. Taking an example, age greater than 60 may indicate that the font sizes on a mobile map should be enlarged, and the type of a mobile device being ‘mobile_phone’ may indicate that the font size should be reduced. Suppose now that we have a context C_1=(o,65,o,mobile_phone,o). In this context the outcome would be that the text fonts are enlarged since the age parameter is considered to be more important than the device parameter. The purpose of these context assignments is then to control the map contents and visualisation in the adapted map to be delivered from the map service. An example of seasonal maps for teenagers on a PDA is shown in Figure 2. In the figure, three context parameters control the map: p_{TIME}∈{“season:winter”,”season:summer”}, p_{AGE} =”teenager”, p_{DEVICE} =”PDA”. The context parameter p_{AGE} reflects to the choice of the style of the icons (for teenagers a comics style used). Depending on the context parameter value for p_{TIME}, the delivered map has different contents and colours defined by the map specification modifiers.

Fig. 2. Adaptive PDA maps for different seasons for the teenager user group.

3.2 Map specifications

When using the mobile map application, the user performs activities to fulfil the needs she/he has in the current context. Map applications do not necessarily need to 'know' all possible activities that users can perform (Meng et. al., 2005). Instead, activities can be split into goals, sub-goals and actions that need to be supported by the service. In the presented approach the application is able to aid the user to reach the goals by implementing so called map specifications.

A *map specification* describes all the properties of a map for a given usage situation, e.g. the parameters for the level of details of the map and map generalisation, and settings for the map layout, symbols, contents, which of the feature classes should be displayed, in which colour, what should be the width and style of the lines, what kind of map icons are used in the adapted map, etc. While the context parameters describe the need for a certain type of map adaptation, a map specification is a solution for this need. A map specification can be related to many contexts, i.e. the same kind of map can be useful in various usage situations.

Usually, a map specification does not contain any user activities directly. Since the context is assumed to be constant in a certain moment of time and the properties of a map are generated only from the map specifications, the user actions are always supposed to change the context. In this way, the user's usage history can be seen as a discrete set of contexts, and user activities are changes from one context to another.

In a real map specification knowledge base there can easily be dozens of different map specifications. Patterns can often be found among map specifications so that two maps can almost be similar, except for one particular map specification modifier. Therefore, in order to be able to define similar map specifications more easily, we divide them into so-called map specification modifiers. We combine these specification modifiers by using a multiple inheritance-based approach, and thus it is possible to both decrease the amount of manual work and create more generic map specifications. We denote the specification modifier as X_P, defined by the context parameters:

$$X_P = (x_1, \ldots, x_m), \text{ where } x_i \text{ are properties of the map} \qquad (2)$$

There is no need for every map specification modifier to have all properties defined. We denote properties that are not defined as *o*. For example, a map specification modifier could be

$$X_{(o,\text{handheld},o,25,o)} = (x_{1a}, o, o, x_{4a}, o), \text{ where } x_{ia} \text{ are some map properties} \qquad (3)$$

Map specification M is the result of the interpretation of a set of map specification modifiers by function h:

$$M = h(\{X_1,\ldots,X_n\}) = (x_1,\ldots,x_m) \qquad (4)$$

A map specification cannot have empty values in its property set.

3.3 Implemented algorithm

The interpretation function h that was implemented into the GiMoDig prototype service uses a method that could be seen as a kind of multiple inheritance. The algorithm applies map specification modifiers ranging from the most common ones to more specific ones. All properties that are present in the 'subclass' will overwrite existing property definitions applied from 'super classes'. This approach makes it possible to have very generic specification modifiers that are useful for sharing common map properties in many map specifications. The following is the pseudo-algorithm we have used to apply map specification $X_{(c1,\ldots,cn)}$:

1. Make a set of map specification modifiers that are 'subclasses' of X, i.e. $D = \{ d_P \mid P=(p_0,\ldots,p_n)$, where $\forall i$ ($p_i=c_i$ or p_i=o) }.
2. Let $i:=1$.
3. Apply map specification modifier d_P from D that has $p_i=c_i$. If there are many, apply the one that has the smallest ordering value, which is the number of the last (i.e. least important) parameter that has a value other than o. If there are no such modifiers, go to step 5.
4. Carry out operation X:=inherit(X, d_P). Remove d_P from D.
5. Increase i by 1. If $i \leq n$ go to step 3, otherwise stop here.

Operation inherit(X,Y) simply overwrites all the map properties in $X=(x_1,\ldots,x_m)$ that are defined in $Y=(y_1,\ldots,y_m)$. More formally:

$$\text{inherit}(X, Y) = (\, a_1,\ldots,a_m \mid \forall i \in [0,\ldots,m],\ a_i = x_i \ \text{if}\ y_i = o \ \ ; \ \ y_i \ \text{otherwise}\,) \quad (5)$$

Two examples of how the implemented algorithm works are presented in Table 2. The map specifications M_1 and M_2 represent specifications generated by the algorithm in given contexts and with specification modifiers W, X, Y and Z. In the following, the algorithm generating the map specification M_1 is studied:

D_{M1} = {W ,Y ,Z} and *i*:=0: Apply *W* and do inherit(M_1,W) => M_1=(w_1,w_2,w_3,w_4,w_5). Remove W from D_{M1}=>{Y, Z} and increase *i* (*i*=1). Apply *Z* and carry out inherit(M_1,Z) => M_1=(w_1,w_2,w_3,z_4,z_5). Remove Z from D_{M1}=>{Y} and increase *i (i*=2). Apply *Y* and carry out inherit(M_1,Y) => M_1=(w_1,w_2,w_3,y_4,y_5). Remove Y from D_{M1}=>{ }and stop.

M_1 has all context parameters defined and modifiers W, Z and Y are applied. Z is applied before Y because Z has a smaller ordering value. The result contains properties only from W and Y because applying Y would overwrite all properties modified in Z. M_2 does not have all context parameters defined, but still it is able to produce a complete set of map properties. M_2 will be constructed from modifiers W, X and Z.

Table 2. Two examples of map property sets created by the algorithm using context parameters and map specification modifiers.

Map spec. modifier	Context	Ordering value	Map properties
W	(k,o,o,o,o)	1	(w_1,w_2,w_3,w_4,w_5)
X	(o,o,n,q,o)	4	(x_1,x_2,o ,o ,x_5)
Y	(o,l,o,m,o)	4	(o ,o ,o ,y_4,y_5)
Z	(o,l,n,o,o)	3	(o ,o ,o ,z_4,z_5)
Specification M_1	(k,l,n,m,p)	-	(w_1,w_2,w_3,y_4,y_5)
Specification M_2	(k,l,n,q,o)	-	(x_1,x_2,w_3,z_4,x_5)

3.4 Meeting the requirements of context modelling

Gross and Klemke (2003) studied general properties of context models in a field of information retrieval. They identified ten requirements that every context model should fulfil:

1. The model identifies all relevant contextual dimensions.
2. The context knowledge is associated to information.
3. Context information is used in queries to community information.
4. Context-based and content-based retrieval of information can be done separately.
5. Context is recognised automatically.
6. Partially matching contexts are used to give the user a coherent view to information.
7. The model has a dynamic ranking of important contextual dimensions.
8. The notification with relevant events considers user preferences.

9. The modelling and maintaining context model clearly increases the working efficiency.
10. Recognising the current context is done in reasonable time.

The presented approach clearly fulfils most of these requirements. Since the approach leaves content of context parameters (i.e. dimensions) open, the responsibility for fulfilling the first requirement is left for the third party developers. With the aid of empty values *o* and ranking the context parameters by their order, the MSKB always produces the ‘best’ possible map in every context and therefore the requirements 6 and 7 are fulfilled. The way of separating the MSKB from the other processing (such as data integration, storage, visualisation, etc.) fulfils requirement 4. The requirements 3 and 8 are out of the scope of the presented research.

4 Map Specification Tool

In order to provide expert users, such as third-party service developers with the possibility of easily creating, updating and modifying map specifications into the knowledge base for their own needs, a sophisticated web-based Map Specification Tool was implemented. Connecting the given context parameters to the map properties allows an easy definition of a wide range of map layouts, as seen in Figure 3.

The Map Specification Tool is implemented as a Java servlet, operating fully on a web browser. Map specification modifiers are saved in plain text files on the server disk, so that they can be easily copied to different services. Alternatively, more effective storage, for example a relational database or XML-files, could be used. The Map Specification Tool is connected directly to the Value-Added Service layer (Figure 1) so that the changes are immediately visible in the real service. Thus, Map Specification Tool makes it possible to control the following properties of the map types in real time (Sarjakoski and Sarjakoski, 2004; Sarjakoski et al., 2005):

- Topographic feature classes to be shown on the map
- Poits of Interest/ Areas of Interest/ Lines of Interest (POI/LOI/AOI) data to be shown on the top of the topographic data
- LoD of the map and generalisation operators, and their parameters that are to be executed on the topographic features
- Other visualisation operators to be executed (for example icon placement)
- Other map visualisation (for example colours, line widths)

Fig. 3. Screen shots of the graphical user interface for the Map Specification Tool Editor. The service administrator selects a prototype map by the context parameters (left) and then defines the specification modifiers (right). The more context parameters are left unassigned (n/a), the less specialised any given prototype map is. Any part of the specification modifier can be derived or inherited from the less specialised prototype maps.

Once the design issues have been carefully resolved, the use of Map Specification Tool can be divided into three phases (Figure 4):

- Create a complete property set to be used as a default map specification modifier. Doing this guarantees that all the combinations of context parameter values produce a complete map.
- Create all map specification modifiers by moving from the most generic to the most specific.
- Verify and experiment with real contexts and fix/create new modifiers if needed.

While the map specification modifiers are combined by using a multiple inheritance-based approach, it is possible to both decrease the amount of manual work that needs to be done and to create more generic map specifications which makes it easy to use the approach for creating various applications. The importance of meticulous design in creation of a new map specification knowledge base cannot be overestimated. The planned structure of the knowledge base and selection of context parameters that will be used have a great impact on how easy or difficult it is to create specifications.

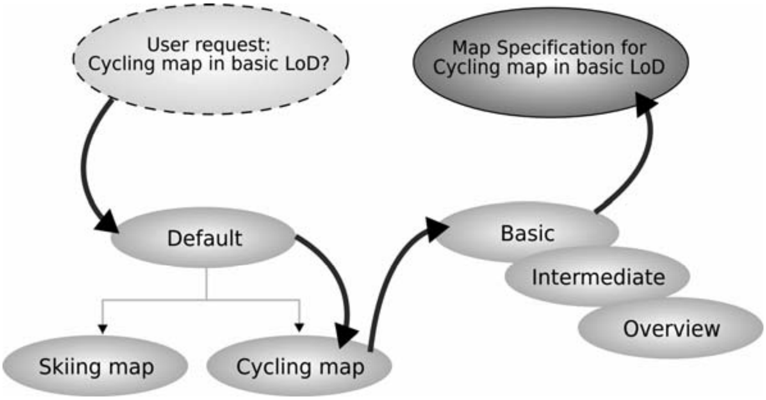


Fig. 4. An example of the applying process of map specification modifiers when p_{USE_CASE} =”cycling” and p_{LOD} =”basic”.

Figure 5 shows an example of different kinds of adapted cycling maps for different types of devices. The maps in the figure have been delivered in real time from the GiMoDig service based on the varying context parameters. Similar solutions to the presented approach can be found from the literature. For example, Weiser and Zipf (2005) presented a visual editor that can be used to generate OGC’s Styled Layer Descriptions for WMS map services. These layer descriptions could be used in a similar way as the map specifications presented in this paper.

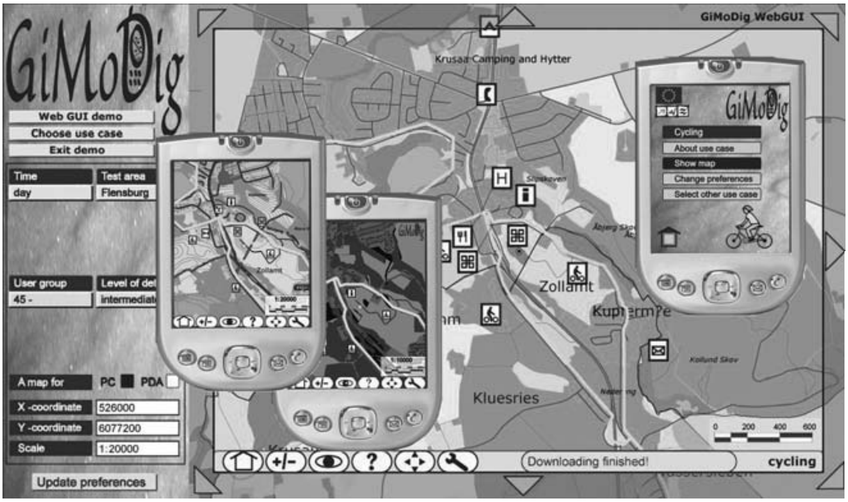


Fig. 5. Different types of cycling maps delivered in real-time by the GiMoDig service: day and night cycling maps for a PDA and a day cycling map for a laptop PC.

5 Concluding remarks

The chapter presented a Map Specification Knowledge Base approach for using and managing different map specifications for the different user needs of mobile maps. Additionally, the web-based Map Specification Tool described here was implemented to provide third-party service developers with the possibility of easily create and modify map specifications for their own applications. The tool provides a flexible way for application developers to elaborate in real time with the map properties and immediately get the response map delivered from the service. This helps the designers more effectively to find the best cartographic visualisation for the current purpose of the mobile map.

The innovative aspect of the MSKB and the map specification approach is that the service can deliver various types of maps, which match the current context parameters and user preferences in real-time. While the map specification modifiers are combined by using a multiple inheritance-based approach, it is possible to both decrease the amount of manual work that needs to be done and to create more generic map specifications which makes it easy to use the approach for creating various applications. The knowledge-based approach formally described here is implemented in the GiMoDig prototype to deliver adaptive mobile maps for different types of users in different usage situations. Development of the approach will continue towards an operational system for adaptive maps. Such maps will be an essential part of the LBSs, of the not too distant future.

Acknowledgements

This research was part of the Geospatial Info-mobility Service by Real-time Data-integration and Generalisation (GiMoDig) project, IST-2000-30090, funded by the European Union through the Information Society Technologies (IST) programme.

References

Baus, J., Kray, C. and Krüger, A., 2001, Visualization and route descriptions in a resource-adaptive navigation aid. *Cognitive Processing,* Vol., 2, No. 2-3, pp. 323-345.

Buttenfield, B.P. and McMaster R.B., (eds.), 1991, *Map Generalization: Making Rules for Knowledge Representation.* Longman Group, London.

Dey, A. K., 2001, Understanding and Using Context. *Personal and Ubiquitous Computing,* Vol. 5, No. 4–7.

Dey, A. and G., Abowd, 1999. Towards a Better Understanding of Context and Context-Awareness, Technical Report, GIT-GVU-99-22, Georgia Institute of Technology, Atlanta (GA).

Forrest, D., 1993, Expert systems and cartographic design, *Cartographic Journal,* Vol. 30, No. 2, 143-147.

GiMoDig, 2005, Geospatial info-mobility service by real-time data-integration and generalisation. IST-2000-30090. Online: http://gimodig.fgi.fi/ (26/09/06).

Goulimis, E., D. and Spanaki, M. and L. Tsoulos, 2003. Context-based cartographic display on mobile devices. In: G. Gartner (ed.), Location Based Services & TeleCartography, Proceedings of the Symposium 2004, Vienna University of Technology, January 28-29, 2004, Vienna,

Gross, T. and Klemke, R., 2003. Context Modelling for Information Retrieval – Requirements and Approaches. *IADIS International Journal on WWW/Internet 1,* Vol. 1, June 2003, pp. 29-42.

Gross, T. and Specht, M., 2001. Awareness in Context-Aware Information Systems.. In Oberquelle, H., Oppermann, R. and J. Krause, (eds.), *Mensch & Computer – 1st Fachuebergreifende Konferenz*, Teubner, pp. 173-182.

Lehto, L. and Sarjakoski, T., 2005, XML in Service Architectures for Mobile Cartographic Applications. In: Meng, L., Zipf, A. and Reichenbacher T., (eds.), *Map-based Mobile Services, Theories, Methods and Implementations,* Springer Berlin Heidelberg New York, pp. 173-192.

Meng, L., Zipf, A. and Reichenbacher T., (eds.), 2005, *Map-based Mobile Services, Theories, Methods and Implementations,* Springer Berlin Heidelberg New York.

Müller, J. C. and Zeshen, W.,1990. A knowledge based system for cartographic symbol design. *The Cartographic Journal,* Vol. 27, No.1, pp. 24-30.

Nivala, A.-M. and Sarjakoski, L. T., 2003, Need for Context-Aware Topographic Maps in Mobile Devices. In: Virrantaus, K. and Tveite H., (eds.), *ScanGIS'2003, Proceedings of the 9th Scandinavian Research Conference on Geographical Information Science,* June 4-6, 2003, Espoo, Finland, pp. 15-29.

Nivala, A-M. and L.T. Sarjakoski, 2005. Adapting Map Symbols for Mobile Users. *Proc. of the International Cartographic Conference 2005: Mapping Approaches into a Changing World*, July 9-16, 2005, A Coruña, Spain, CD-ROM: Theme 12: Internet Location-Based Services, Mobile Mapping and Navigation Systems, Session 5.

Oppermann, R., (ed.), 1994, *Adaptive User Support: Ergonomic Design of Manually and Automatically Adaptable Software, Computers, Cognition, and Work.* Hillsdale (NJ): Lawrence Erlbaum Associates.

Reichenbacher, T., 2001, The World in Your Pocket – Towards a Mobile Cartography. *Proceedings of the 20th International Cartographic Conference,* Beijing, China, pp. 2514-2521.

Reichenbacher, T., 2004, Mobile Cartography – Adaptive Visualisation of Geographic Information on Mobile Devices. Dissertation, Department of Cartography, Technische Universität München, München: Verlag Dr. Hut.

Sarjakoski, L.T and Nivala, A.-M., 2005, Adaptation to Context – A Way to Improve the Usability of Mobile Maps, In: Meng, L., Zipf, A. and Reichenbacher T., (eds.), *Map-based Mobile Services, Theories, Methods and Implementations,* Springer Berlin Heidelberg New York, pp. 107-123.

Sarjakoski, L. T., Koivula, T. and T. Sarjakoski, 2005. Knowledge-based map adaptation for mobile map services. In: Gartner, G. (Ed.), Proceedings of the Symposium ”Location Based Services and TeleCartography”, Vienna, Austria, November 28-30, 2005, *Geowissenschaftliche Mitteilungen,* Technische Universität Wien, Heft Nr. 74, 2005, pp. 33-40.

Sarjakoski, L.T., Nivala, A-M. and Hämäläinen, M., 2003, Improving the Usability of Mobile Maps by Means of Adaption. In: G. Gartner (ed.), *Location Based Services & TeleCartography, Proceedings of the Symposium 2004*, Vienna University of Technology, 2004, Vienna, pp. 79-84.

Sarjakoski, L., T. and Sarjakoski, T., 2004. A Use Case Based Mobile GI Service with Embedded Map Generalisation. Papers of the ICA Workshop on Generalisation and Multiple Representation, 2004, Leicester. Online: http://ica.ign.fr/Leicester/paper/Sarjakoski-v2-ICAWorkshop.pdf (26/09/06).

Sarjakoski T. and Sarjakoski, L. T., 2005. The GiMoDig public final report. GiMoDig-project, IST-2000-30090, Deliverable D1.2.31, Public EC report. Online: http://gimodig.fgi.fi/deliverables.php, (26/09/06).

Sarjakoski, T., Sester, M., Sarjakoski, L.T., Harrie, L., Hampe, M., Lehto, L. and T. Koivula, 2005. Web generalisation services in GiMoDig - towards a standardised service for real-time generalisation. In: Toppen, F. & M. Painho, (eds.), *Conference proceedings, AGILE 2005, 8th Conference on Geographic Information Science,* Estoril, Portugal, May 26-28, 2005, pp. 509-518.

Stefanakis, K. and L. Tsoulos, 2005. Structure and development of a knowledge base for cartographic composition. *Proc. of the International Cartographic Conference 2005: Mapping Approaches into a Changing World*, July 9-16, 2005, A Coruña, Spain, CD-ROM: Theme 3: Map Design and Production.

Tomai, E. and M. Kavouras, 2005. Context in Geographic Knowledge Representation, *Proceedings of the II International conference & exhibition on geographic information, GIS PLANET,* Estoril, Portugal, May 30 - June 2, 2005, CD-ROM.

Weiser, A. and A. Zipf, 2005. A visual editor for generating OGC Styled Layer Description (SLD) files for automating configuration of Web Map Server (WMS) maps. In: Gartner, G. (Ed.), Proceedings of the Symposium ”Location Based Services and TeleCartography”, Vienna, Austria, November 28-30, 2005, *Geowissenschaftliche Mitteilungen,* Technische Universität Wien, Heft Nr. 74, 2005.

A visual editor for OGC SLD files for automating the configuration of WMS and mobile map applications

Albrecht Weiser, Alexander Zipf

i3mainz - Institute for Spatial Information and Surveying Technology
Department of Geoinformatics and Surveying
University of Applied Sciences FH Mainz, Germany
zipf@geoinform.fh-mainz.de

Abstract

The OGC Styled Layer Descriptor (SLD) specification [OGC 2003] allows describing the design and contents of a map – in particular a map served by a SLD-WMS – in a formal way through the use of a strict XML schema. The specification is getting more and more used and supported by several implementations of Web Map Servers. But SLD still has some weaknesses – especially for use in LBS, see e.g. [Brinkhoff 2004]. We present here the first implementation of a mobile map application (mobile city guide running on a PDA), that uses the SLD specification as configuration of the mobile map. On the other hand in LBS or UbiGIS maps need to be generated dynamically taking a lot of further factors into account – in particular also attributes describing the current user, the situation and general context [Zipf 2002, Reichenbacher 2004, Meng et al. 2004, Zipf and Jöst 2006]. In order to use this kind of parameter and information in an automated map service on the internet or on mobile devices we need also formal representations for this type of information – e.g. also as XML files. Then we need to combine the information from this context models and the base map which is represented through a SLD file to automatically generate adapted maps according to those parameters. The

result is again an SLD file describing the adapted map, which then can be used within a request to an SLD-enabled WMS or a mobile application that supports SLD. So the first task is to generate the SLD for the base map. In order to automate the task of designing maps for Web Map Servers or mobile map applications we introduce a free software tool called ArcMap2SLD-Generator (http://arcmap2sld.geoinform.fh-mainz.de/). This tool is helpful as it allows generating a valid SLD-file from the design of a map within ESRI ArcMap. The resulting SLD in turn can act as a base for further modifications through user or context models in order to generate user-adapted map representations as SLD [Zipf 2005] in a standard-conformant way.

1 The OGC SLD Specification - a formal representation of maps

The OGC Styled Layer Descriptor Specification [OGC 2002] defines a XML schema to describe the appearance of the layers of a map. A SLD document is a XML file that can be validated against this model. SLDs are getting more popular in web mapping applications with the growing availability for SLD support in WMS. But until recently these SLDs are more or less hand-made or application specific. But as SLDs provide the means to specify the look of maps in a domain and vendor-neutral way, it is a good choice for a formal representation of maps in general.

The question is now how to go beyond these hand-made SLDs and generate these in an automated way using open standards predominantly. This is a technical question and can be solved using a software tool we have developed. This tool is a SLD generator for ESRI ArcMap maps. This can be used not only in WMS requests or for configuring WMSs (as in the case of the deegree WMS), but also when producing vector based maps in SVG. Merdes et al. [2005] present an example of using SLD and GML from WFS requests with a cascade of several XSL- transformations in order to generate a SVG map from GML and SLD. It has been shown how this can be extended to include user and context models in order to generated user-adapted SLD files in a generic way (Zipf 2005). First we will briefly introduce our application that generates SLD files from existing maps in commercial desktop GIS. Then we explain the results of an evaluation of SLD support by two major WMS implementations (UMN and ArcIMS) and afterwards we introduce the concept of SLD for mobile maps and present a first implementation of SLD support within mobile map applications for PDAs.

2 The ArcMap2SLD Converter

In this section we want to explain the functionality and the benefits of the ArcMap2SLD-converter.

The first question is: What is this tool designed for at all? The answer is that it was developed mainly for supporting the setup of a WMS. When setting up a WMS it is much work to write the code for the symbolization prescription of the configuration file. If you want to set up a large project with various layers and many classes, by far the most effort is to code the hundreds (or even thousands) of lines of the symbolization prescription. So the question arises how this can be automated in order to minimize such work? The optimal case would be if one could reuse the effort and work, that was already invested in setting up a project and designing a map with a common desktop GIS. This would allow to take the power of desktop GIS supporting the symbolization process also for Web-based maps. If it would be possible to analyze the symbolization of a map in a desktop GIS and automatically transform the information into a SLD, the most time-consuming activity in setting up a WMS could be saved.

In order to realize and evaluate this approach we developed a tool for converting the symbolization of an ArcMap project into the SLD format. This so called ArcMap2SLD-converter analyzes the symbolization of an ArcMap-project by stepping through each layer of the map and each class of each layer. Then it stores the symbol properties along with further properties of each layer as well as the overall project in an internal data structure. This allows a very dynamic analysis also of the complex symbol objects of ArcMap such as MultiLayer-symbols. After finishing the analysis, all data stored in these data structures can be transformed into SLD and exported into a SLD-file. It is also possible to check the validity of the resulting SLD by integrating the relevant SLD schema-files from the OGC. The SLD can then be referenced by the WMS in the reqest-URL. The following example shows this approach:

http://127.0.0.1/mapserver?SERVICE=WMS&VERSION=1.1.1&
REQUEST=GetMap&Layers=Guek300Poly&STYLES=&SLD=http://127.0.0.
1/SLD/GUEK300.sld&BBOX=3377854,5471076,3592188,5728258

The reason to take the ArcGIS system from ESRI as starting point was its widespread distribution as market leader in the GIS market and the abundance of existing maps designed already with ArcMap. For this reason a large number of GIS user can participate the benefits of the ArcMap2SLD-converter. Some of the technical design parameters of the ArcMap2SLD-Converter are mentioned below:

- The application was developed in Visual Basic.NET
- It uses the ArcObjects framework from ESRI for accessing ArcGIS-functionalities
- The navigation in the SLD-file is performed by XPath (XML Path Language)
- Syntactic changes of the SLD specification can be adjusted in a configuration file

2.1 Implementing the ArcMap2SLD-converter

The many possibilities how to classify maps are supported by ArcObjects through a comprehensive internal data model. This model is the ESRI “ArcObjects” class library (Burke 2003). The interfaces of the library present the functionality of the respective classes. This allows to use the properties and methods of the library through any programming language supporting the Microsoft .Net-platform. According to the options how to classify maps in ArcMap the following Renderer-types are supported in Arc Objects.

1. *SimpleRenderer* ≈ Single symbol (Features)
2. *UniqueValueRenderer* ≈ Unique values (Categories)
3. *ClassBreaksRenderer* ≈ Graduated colors, Graduated symbols (Quantities)
4. *ProportionalSymbolRenderer* ≈ Proportional symbols (Quantities)
5. *DotDensityRenderer* ≈ Dot density (Quantities)
6. *ChartRenderer* ≈ Charts
7. *BiUniqueValueRenderer* ≈ Quantity by Category (Multiple Attributes)

ArcObjects supports also five main symbol types with several symbol subtypes. The latter are TextSymbol, MarkerSymbol, LineSymbol, FillSymbol and 3DchartSymbol.

The converter realized currently supports the most important means of thematic classification within ArcMap:

- *Features* (Single Symbol)
- *Categories* (Unique values; Unique values, many fields)
- *Quantities* (Graduated colors)

Further *Marker*-, *Line*- and *Polygon*-features are also supported by the application as well as text-symbols. Even multilayer-symbols are supported for these features. Multilayer-symbols are able to depict complex symbolization like the filling of an area with repeating patterns. MultiLayer-Symbols can aggregate a number of symbols of the same main symbol type. Each symbol can be aggregated from several symbols, e.g. polygon symbols can include line symbols or marker symbols.

In order to use the converter we need to have ArcGIS installed on a running ArcMap-Session with the project open that includes the map to be transformed. The converter application is designed to minimize user interactions and to achieve a maximum of automatization.

- All symbols present in the current ArcMap map are exported into internal software structures
- all render-types of ArcObjects have been implemented that can be used in SLD currently. These are: UniqueValueRenderer (incl. many fields) / SimpleRenderer / ClassBreaksRenderer
- Not yet fully implemented are: ChartRenderer/ DotDensityRenderer / ProportionalSymbolRenderer / ScaleDependentRenderer

The following visualization properties are already implemented for the individual feature classes:

- *Point-Feature*: colour, dot-size, angle of rotation
- *Line-Feature*: colour, line width
- *Polygon-Feature*:
  - *One colour area filling*: area colour, line colour and line width of border
  - *Point filling*: Point colour , point size, line colour and line width of border
  - *Cross hatch*: Line colour and line width of the hatching, vertical and diagonal angle of hatching, line colour and line width of border

The visualization of *Multilayer-symbols* as two-layer symbols has also been implemented. We support only two layers in Multilayer-symbols so far as the current WMS implementations such as UMN are not able to process more than two-layers in Multilayer- symbols at the moment. In order to label the symbols in the legend for every rule of a symbol the title-element is included. Line hatching as well as the use of bitmaps for area fillings are currently not yet supported.

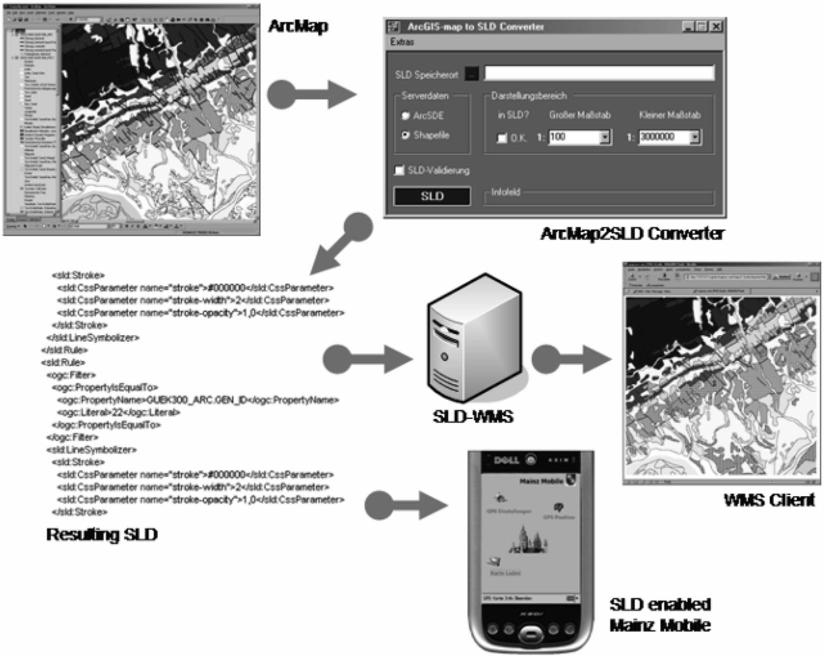


Fig. 1. Process of designing a map in ArcMap and exporting it to a SLD document. The latter can be used to configure WMS servers or MainzMobile or to issue user-specific requests to SLD Web Map Services.

3 Evaluating the SLD support of WMS servers

In order to evaluate the possibilities of representing the visual appearance of maps with SLD on current WMS servers we conducted several tests using the generated SLDs. The goal was to answer the following four main questions:

1. How good is the support of the evaluated test servers (UMN, ArcIMS) for the SLD specificatin?
2. What possibilites to style maps does the current specification offer? The main aspects looked at included:
- Symbolizer (Point-, Line-, PolygonSymbolizer)
- Complex area fillings with hatchings and point fillings
- Classifications using filter and logical operators
3. What further extensions to the converter are usefull?

4. What configurations are needed in order to make the evaluated test server visualise the SLD the best way (according to their respective possibilities)?

One of the results of applying the generated SLD file to the UMN WMS can be seen in figure 2, which compares the map to the original ArcMap display. One can recognize from figure 2 that it is still not possible to achieve 100% the same look using the WMS on the one hand and the original ArcMap map on the other, but a very good compromise can be achieved. Where does the different look come from? The fact is, that SLD is still standard in development. Its contents increases step by step during the suggestions of its members. For example until now SLD still has no possibilities to generate a vector-based hatching for polygons, because the SLD-element *Graphic* of the SLD-element *PolygonSymbolizer* has only two successor-elements: *Mark* and *ExternalGraphic*. This implies that SLD can only generate vector-based area filling patterns with marker symbols because a successor-element *Line* is missing at present. Among others this is a reason for the different looks of the source map and the resulting map. Nevertheless, SLD is sufficiently extensive to portray a complex symbolized map to distinguish all containing classes in a satisfying way.

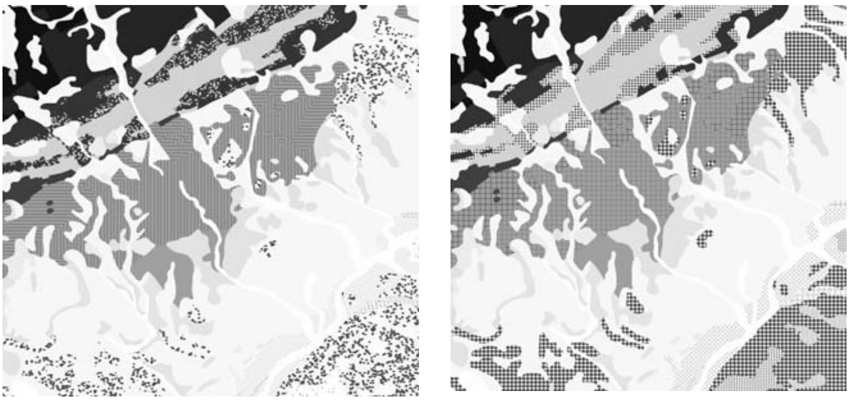


Fig. 2. Comparison of the resulting map (detail) from the dynamically generated SLD description renderend by UMN Mapserver by SLD (right side) and the original map in ArcMap (left side). Example: Geologische Übersichtskarte 1:300.000, GÜK 300, Hessisches Landesamt für Umwelt und Geologie (HLUG) (Geological Map 1:300.000)

Currently a range of map server implementations still offer even less support for the different parameters defined in SLD then the ones we evaluated. Therefore maps generated using such WMS implementations even may look more different. We assume that it is only a matter of time until this varying support of the SLD specification has reached a more stable situation. Then the different render engines (the different WMS implementations) should generate quite similar maps from the same SLD configuration. In order to achieve this aim also the SLD specification does need further extensions in order to clarify how to represent a range of symbolization issues. In figure 3 we see the result of ArcIMS with WMS-Connector displaying the same map styled be the same SLD file. It looks even more different from the ArcMap original than the UMN version. It seems that the version 1.0 of the WMS-Connector of ArcMap does not support all visualization possibilities that ArcIMS itself would support leading to an insufficient result. When we visualize the map with ArcIMS without SLD, but instead with the ESRI-proprietary ArcXML configuration, one can see that the accordance with the original ArcMap version (figure 2 left) is slightly higher.

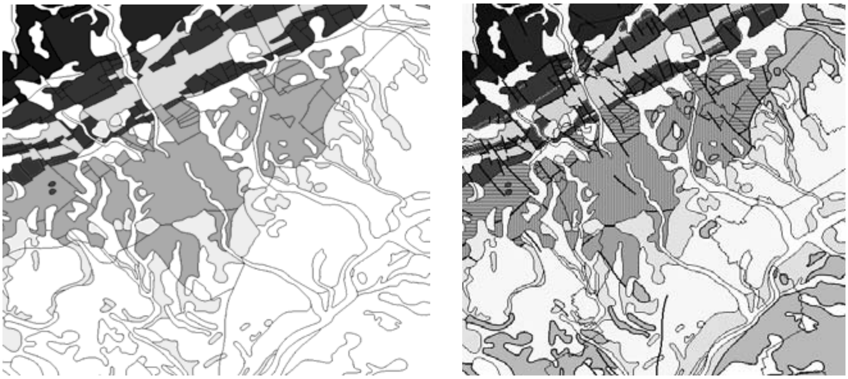


Fig. 3. Comparison of the resulting map (detail) from the dynamically generated SLD description rendered by ArcIMS Mapserver by SLD (with WMS connector) (left side) and the proprietary ArcIMS map using ArcXML without use of open standards as SLD (right side). Example: Geologische Übersichtskarte 1:300.000, GÜK 300, Hessisches Landesamt für Umwelt und Geologie (HLUG) (Geological Map 1:300.000)

Figure 4 shows as result a comparison of the support for each of the SLD elements that have been tested using both UMN and ArcIMS.

SLD-element	Support UMN	Support ArcIMS	SLD-attribute	Support UMN	Support ArcIMS
NamedLayer	+	+			
UserLayer	-	-			
NamedStyle	+	+			
UserStyle	+	+			
FeatureTypeStyle	+	+			
Rules	+	+			
Min-, MaxScaleDe-nominator	+	+			
Filter	+	+			
ElseFilter	+	+			
PropertyIsEqualTo	+	+			
PropertyIsBetween	+	+			
AND	+	+			
PointSymbolizer	+	+			
Fill	+	-			
			Fill*	+	-
			fill-opacity *	-	-
Size	+	-			
Mark	+	-			
WellKnownName	+	-			
Rotation	-	-			
LineSymbolizer	+	+			
Stroke	+	+			
			Stroke*	+	-
			stroke-width*	+	-
			stroke-opacity*	-	-
			stroke-dasharray*	+	-
GraphicStroke	+	-	(s. Stroke)		
PolygonSymbolizer	+	+			
Fill	+	+			
			fill*	+	+
			fill-opacity*	-	-
GraphicFill	+	-			
Graphic	+	-			
ExternalGraphic	+	-			

Fig. 4. Tested SLD elements and attributes tested on UMN 4.4.2, 4.6.0 beta2 and ArcIMS 4.0.1 (with OGC WMS Connector 1.0) for the GÜK 300 geologic overview map.
* CSS parameter

4 SLDs for configuring mobile maps

Many commercial applications for PalmOS / Pocket PC PDAs exist that display the current position on a map in combination with a GPS receiver and allow stand-alone routing. Most of these applications are targeted towards car navigation, but they can also be used for pedestrian navigation in a city in a limited way. An increasing number of commercial car navigation systems also offer a 3D view, but this usually does not deliver a real 3D representation of the area, but only a perspective view of the 2D map. They usually do not make use of open OGC standards, as the SLD specification or the W3DS discussion draft. Also mobile GIS applications like ArcPad etc. do not offer the possibility to use SLD for configuring map styles.

Our PDA application - MainzMobile3D - is a mobile 2D and 3D map application for PocketPC supporting both SLD and the W3DS specifications (Fischer et al. 2006). It is running locally on the PDA itself and has been developed in C# on the .Net Compact Framework 2. Currently the following functionalities have been realized among others:

- positioning by means of GPS
- internal geometry model based on OGC Simple Feature Specification
- 2D map display using multiple generalization levels
- 2D map functions (zoom, pan etc.)
- search and visualization of POIs
- search function for object information / touristic information
- distance measurement tool
- GPS configuration & display (Skyplot, Skyview, coordinates etc..)
- 3D scene information as on-line 3D Web-Client based on the OGC Web3D service Specification Discussion paper (Quadt and Kolbe 2005)
- tour planning (A* algorithm)(Köhler 2006)
- ESRI *.shp files as local data source for geodata
- Spatial index
- internal database for POIs
- focus maps prototype (Schuler 2006) based on ideas by Zipf and Richter (2002)
- SLD based configuration of the map display

The latter functionality shall be explained in a little more detail in the following section. To our knowledge MainzMobile3D is the first and only PDA-based mobile map service that supports SLD for configuration of the maps generated directly on the mobile device.

The first version of MainzMobile3D had a simple proprietary map configuration based on XML (Fischer 2006). It already supported scale-dependent map configurations and symbols. Lindemann (2006) extended that approach by implementing SLD-support of MainzMobile3D. This required mainly some changes of the XML data structure used for the configuration. On the other hand he also extended the ArcMap2SLD-Converter to become an authoring tool for the 2D maps in the Mainz Mobile3D application by extending the SLD support (better support for labelling) on the one hand and on the other by developing possibilities to feed the XML-based MainzMobile3D database with multimedia descriptions of points of interest (POI).

The following figures show example screenshots of the authoring tool based on the ArcMap2SLD-converter. As this tool is based on open standards like SLD, it is now possible to apply the MainzMobile3D software much easier to other cities or regions as the content both regarding map design as well as the multimedia POI database can be authored with GUI-based user-friendly tool support. This eliminates the need to manually write XML configuration files.

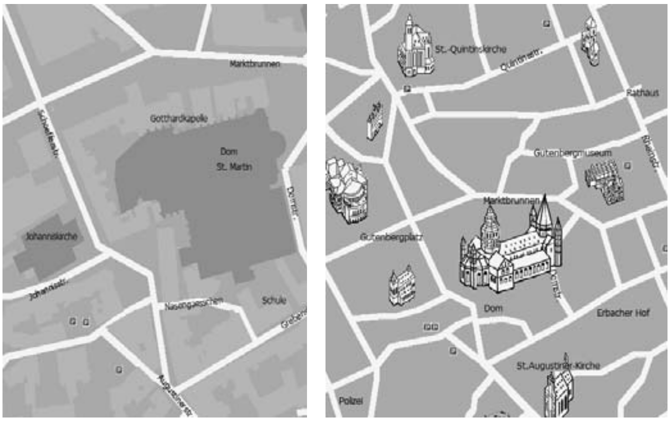


Fig. 5. Scale-dependent rendering through SLD-based 2D-map configuration in MainzMobile3D

Fig. 6. Start-Window of the Authoring-Tool

Fig. 7. Formular for inserting descriptions of POIs

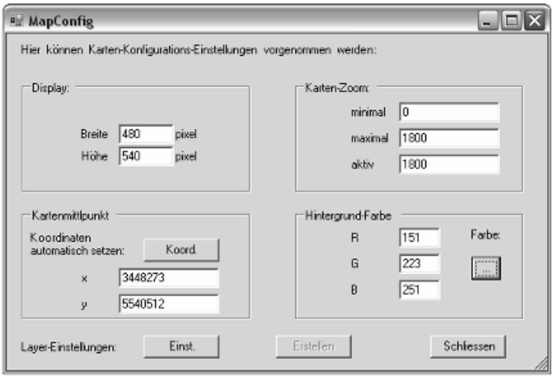


Fig. 8. “MapConfig” -Window of the Authoring-Tool

5 Summary and Outlook

SLD is still a maturing technology. Originally designed in order to allow clients to specify the look of maps served by WMS servers we have shown how it can also be used for configuring WMS as well as mobile map applications running locally on PDAs. In particular the latter option is a completely new way to use the SLD specification as formal map representation directly on mobile devices.

Such innovative use of OGC standards in combination with the supporting tools for generating SLDs and POI databases based on conventinonal GIS is a step towards more interoperable and ubiquitous map services - including desktop GIS, Web Map Services and mobile applications on PDAs or Smartphones.

Further research has shown, that it is possible to derive also SVG-based maps in a completely generic way from standards-based data sources (Merdes et al 2005) and that the technology used for that can even be extended towards the automated generation of user- and context specific SLDs through the incorporation of formal user and context models within a cascade of several XSL- transformation (Zipf 2006). But this is only a solution for the technical side of the problem. How to use such context information in order to generate situation aware and cognitively more adequate maps will still remain an active research question.

Acknowledgements

We like to thank all former and current colleagues for their inspiration and help. We thank the HLUG for their cooperation in developing the ArcMap2SLD-converter and evaluating the SLD support of the WMS using the GÜK300 geologic overview map. We thank the Bureau of Surveying of the city of Mainz for supporting us with test data of Mainz. Parts of the work on the W3D-Service have been carried out with friendly support of the Klaus-Tschira-Foundation (KTS) Heidelberg.

References

Burke R. (2003): Getting to know ArcObjects. Hrsg. ESRI: ESRI Press, 2003

Fischer, M. (2006): Mobiler Stadtführer 3D für PDA. Diplomarbeit. FH Mainz.

Fischer, M., Basanow, Zipf (2006): Mainz Mobile 3D - A PDA based OGC Web 3D Service Client and corresponding server. International Workshop on 3D Geoinformation 2006 (3DGeoInfo'06). Kuala Lumpur. Malaysia.

Brinkhoff, T. (2005): Towards a Declarative Portrayal and Interaction Model for GIS and LBS. Proceedings 8th Conference on Geographic Information Science (AGILE 2005), Estoril, Portugal, 2005, pp. 449-458.

Köhler A. (2006): Tourenplanung mobil auf PDAs. Diplomarbeit. FH Mainz.

Lindemann, M. (2006): Ein Autorenwerkzeug für SLD-basierte Konfiguration mobiler Stadtinformationsdienste. Diplomarbeit. FH Mainz.

Meng, L., Zipf, A. and Reichenbacher, T. (eds.) (2004): Map-based mobile services – Theories, Methods and Implementations. Springer Geosciences. Heidelberg.

Merdes, M., J. Häußler, A. Zipf (2005): GML2GML: Generic and Interoperable Round-Trip Geodata Editing using Standards-Based Knowledge Sources.

AGILE 2005. 8.th AGILE Conf. on GIScience. Assoc. of Geographic Information Laboratories in Europe. Estoril. Portugal.

OGC (2002): Styled Layer Description Implementation Specification, Version 1.0.0.

Quadt und Kolbe (2005): Web 3D Service, OGC Discussion Paper, OGC Reference number: OGC 05-019, Version: 0.3.0

Reichenbacher, T. (2004), Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices, PhD thesis, TU Munich.

Schuler, V. (2006): Ein Fokuskarten-Prototyp für PDAs. Diplomarbeit. FH Mainz.

Weiser, A. (2005): Automatisierte Generierung von Styled Layer Descriptor-Dateien aus ESRI ArcGIS-Projekten zur Publikation mit OGC-konformen Mapservern. Diploma Thesis. University of Applied Sciences. Mainz.

Zipf, A and Jöst, M. (2006): Implementing Adaptive Mobile GI Services based on Ontologies - Examples for pedestrian navigation support. In: CEUS - Computers, Environment and Urban Systems - An International Journal. Special Issue on LBS and UbiGIS. Pegamon Press, Elsevier.

Zipf, A. and Richter, K.F. (2002): Using Focus Maps to Ease Map Reading. Developing Smart Applications for Mobile Devices. In: Künstliche Intelligenz (KI). Special issue: Spatial Cognition. 04/2002. 35-37.

Zipf, A. (2002): User-Adaptive Maps for Location-Based Services (LBS) for Tourism. In: K. Woeber, A. Frew, M. Hitz (eds.), Proc. of the 9th Int. Conf. for Information and Communication Technologies in Tourism, ENTER 2002. Innsbruck, Austria. Springer Computer Science. Heidelberg, Berlin.

Zipf, A. (2005): Using Styled Layer Descriptor (SLD) for the dynamic generation of user- and context-adaptive mobile maps - a technical framework. 5th International Workshop on Web and Wireless Geographical Information Systems (W2GIS). Lausanne, Switzerland. Springer Computer Science. Heidelberg, Berlin.

Towards Orientation-Aware Location Based Mobile Services

Rainer Simon, Harald Kunczier, Hermann Anegg

Telecommunications Research Center Vienna (ftw.)
Donau-City-Str. 1, A-1220 Vienna
{simon, kunczier, anegg}@ftw.at

Abstract

In this chapter we present an approach on how orientation sensors built into mobile devices can enable a new paradigm for mobile service discovery and use. In conjunction with 3D models of urban terrain, mobile devices can act as virtual pointers to services and information anchored at specific locations such as buildings or landmarks. We outline a system architecture that enables this new method of orientation and location sensitive service discovery to be employed. We describe our experiences with a prototype device consisting of a plain, mass market mobile phone and a custom-built shell that houses a magnetic compass and a 2-axis tilt sensor. We conclude by describing our ongoing research project that will apply the concepts and practical technologies described.

1 Introduction

Location based mobile applications are gradually gaining importance in the consumer electronics and telecommunications market. The growing proliferation of increasingly powerful handheld computing devices and the availability of relatively low-cost, embeddable GPS receivers have made navigation applications for PDAs and smartphones a popular product. At the same time, mobile network operators are adopting location based services (LBS) as an increasingly important component of their service portfolio [15]. Typical examples of LBS offered today are direction finding services or yellow pages-like services (e.g. “Where is the next pharmacy?”). Despite the fact that the success of LBS has yet been less than anticipated, user evaluations have shown that the demand for location aware information is high [10]: Since users potentially have access to their mobile device all the time, they expect it to be of particular value for accessing information about unfamiliar environments or locations; when looking for a specific service or in emergency situations; or for accessing the kind of information that may change while they are on the move: Examples are traffic information or train schedules with delay information.

We argue that location aware applications offer a real added benefit to the user – provided that they deliver information that is relevant, easy to find and well-focused. Our work is motivated by the assumption that a more natural and intuitive way of discovering and using location aware services is essential for meeting these criteria. It can not only play a key role for an improved awareness and an increased use of current LBS; it can also open up new possibilities and application areas for future mobile service ideas.

This chapter is organized as follows: In section 2, we introduce the idea of “Point-to-Discover” – our concept of accessing information and mobile services by pointing at geographic locations with a handheld device – and develop a possible system architecture for a Point-to-Discover service platform. In section 3, we address two technical issues that are crucial for the feasibility of our concept: the accuracy limitations of current positioning methods and orientation sensors and the availability of three dimensional environment models. In section 4, we focus on a prototype of a Point-to-Discover-enabled mobile device we developed. We explain the hardware used and point to related work. Finally, in section 5, we present our ongoing research project, in which we will practically apply the technologies and concepts described in this chapter.

2 Point-to-Discover

The advent of 3G technology and more complex mobile devices poses challenges to researchers and designers of mobile applications: The mobile interface is restricted to tiny displays and keypads; information retrieval on the fly is slowed down by bandwidth limitations and network latencies. In all described restrictions, advances can be expected, but whenever interface components increase in size, they impair the mobile devices' mobility. For this reason we envision a new user interface paradigm for how people can discover information on the go and use mobile services in the future, by combining orientation detection with accurate satellite positioning.

Today, users access mobile services and information through WAP menus or search engines, mobile operator portals or by simply entering a URL in their phone's browser. We envisage a more natural method of discovering services and information: With the help of three-dimensional models of urban terrain, services can be 'anchored' virtually at geographic locations. Mobile phones function as pointing devices towards these services. For example, users could access the train schedule, get delay or estimated arrival time information or purchase tickets online by simply pointing their mobile phone towards a train station; or they might participate in a sweepstake by pointing at an advertisement billboard.

An early realization of a similar approach was presented by Wasinger et al [19]: They describe the implementation of a navigation application on a PocketPC PDA. The application uses GPS in conjunction with a magnetic compass and allows the user to indicate an area of interest on a two-dimensional map by pointing the device into a direction in the real world. The focus of their work, however, is primarily on multimodal and speech in- and output in the context of mobile navigation and exploration services.

We suggest to enhance the concept presented by Wasinger et al by adding the third dimension: Using the 3D orientation of the device – measured with a magnetic compass and a tilt sensor – and a 3D environment model rather than a 2D map, the system can determine the user's real perspective and specifically select the services that are available within the user's real field of view. Figure 1 portrays the architecture of a Point-to-Discover system as we envision it. Three domains constitute the system: The client domain consists of the user terminal running the necessary client software ("Point-to-Discover Browser"). The operator domain is formed by the Point-to-Discover service platform. The platform stores geo-referenced meta-data as well as geo-referenced links to content and processes the Point-to-Discover requests. It is important to mention that the actual content itself is not held in the operator domain. The operator domain merely provides the "lookup service" for the content offered by external providers over standard WWW infrastructure (3^{rd} party domain). After the lookup, the client device will access the content directly from the 3^{rd} party services, using the links received from the service platform.

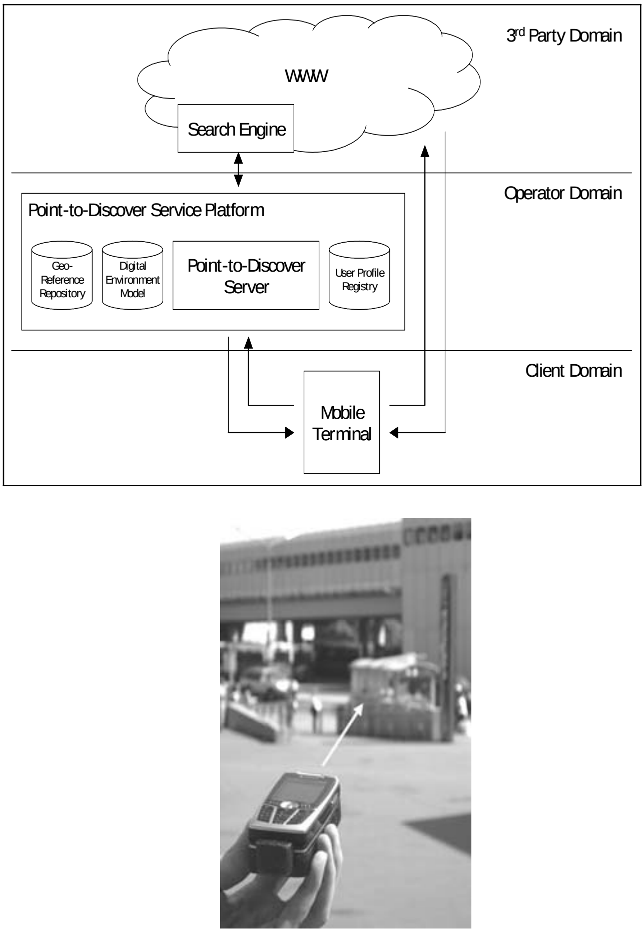


Fig. 1. Point-to-Discover service platform architecture and usage scenario

The Point-to-Discover service platform consists of a digital environment model, the geo-reference repository, the actual Point-to-Discover server and (optionally) a user profile registry. The service platform can – but need not necessarily be-hosted and maintained by a mobile network operator. When the user points at an object in the real world and triggers a request, the mobile phone transmits its position and orientation sensor readings to the Point-to-Discover server. The server queries the digital environment model and identifies the area of interest that is indicated by the pointing ray (defined by the position coordinates and the orientation vector). With the estimated area as a parameter, the server queries the geo-reference repository.

The geo-reference repository is a database that holds a list of geo-referenced links to 3rd party content and services. These links are maintained by the platform operator and therefore allow placement of specially promoted, branded services (such as e.g. the sweepstake linked to the advertisement billboard mentioned in the example above). Optionally, a user profile (stored in the user profile registry) can help to tailor the service selection towards the personalized preferences of the user before the list of service links is returned to the device. In addition to the user profile, the platform might take further external parameters into consideration: Time of day, time of year or the weather at the current location are just a few examples of context parameters that might help to match the service selection better towards the user's current situation.

Additionally, the geo-reference repository also stores geo-referenced meta-data, such as street or landmark names. The Point-to-Discover server can use this meta-data to generate a refined search query for an Internet search engine. That way, the user can obtain information about a landmark (e.g. a particular building of historical relevance) from the World Wide Web without the need to type; and even without the need to know the name of the landmark he or she is pointing at.

3 Technological Requirements

For the successful implementation of a Point-to-Discover service platform as described above, two factors are critical: First, there are considerable accuracy constraints on both positioning and orientation detection that influence the quality of service selection. Second, the Point-to-Discover principle heavily depends on the availability of sufficiently accurate 3D environment models. In this section, we therefore first address the factors that lead to accuracy degradation. Secondly, we want to point out the relationship between our digital environment model and geo-reference repository components and Geographic Information Systems (GIS).

3.1 Selection Quality

Two sources of error influence the service selection process: The positioning error and the error introduced by the orientation sensors. Depending on the positioning method, positioning accuracy in the range of several 10 meters can be achieved in urban environment [14], [17]. If we assume that the Point-to-Discover concept is required to identify an object with a radius r, at a distance d, the overall error is given by

$$e = e_P + d\tan(\alpha) \equiv r$$

with e_p being the positioning error and alpha being the angular error as illustrated in figure 2. If we require to identify an object with a radius r=2.5 meters, at a distance of d=100 meters, the currently deployed terrestrial range based localization technologies are inadequate. The necessary positioning accuracy we would need to achieve is at least 5 meters, with no angular error. Assuming differential code phase GPS, which reaches an accuracy of $e_p \approx 1$ meter, we require a more reasonable angular error of ~1°.

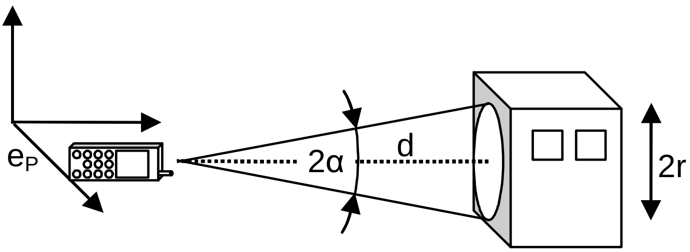


Fig. 2. Angular error

3.2 3D Environment Models

The capabilities of the digital environment model and the geo-reference repository which are part of our Point-to-Discover service platform are closely related to the capabilities of Geographic Information Systems (GIS). Until recently, GIS were mainly restricted to modeling the world in two dimensions due to the high computational effort required for three-dimensional data manipulation and display and, in particular, the complexity involved with gathering three-dimensional environment data. More recent methods such as automatic or semi-automatic building reconstruction from high-resolution aerial imagery and/or LIDAR (LIght Detection And Ranging) scans, however, make gathering of large-scale three-dimensional environment data feasible [9], [13]. Further techniques, for instance combining aerial imagery with ground-level 3D laser scans and/or geo-referenced photographs [7], [11] even aim at automatically reconstructing detailed textured 3D models of urban environment. We argue that these advances in the area of 3D environment modeling indicate that sufficiently accurate large-scale data will indeed be readily available to enable the Point-to-Discover principle. In fact, our platform's digital environment model/geo-reference repository components can be implemented using commercial off-the-shelf 3D GIS, which are already available from a number of vendors and are e.g. being used by mobile network operators for network planning purposes.

4 Orientation-Aware Mobile Devices

Based on an ordinary mass-market Java-enabled mobile phone, we developed a prototype for a Point-to-Discover-enabled client device. The phone has a custom-built shell attached to its back that houses a three-axis tilt compensated compass module purchased from a commercial vendor. The module essentially combines a magnetic compass with a 2-axis tilt sensor on a single chip, mounted to a 2.5 by 4.5 cm printed circuit board. The sensor module is connected to the phone via the serial port.

4.1 Magnetic Compass

The magnetic compass detects the heading of the device by measuring the direction of the earth's magnetic field. Unlike GPS-based methods, which derive the heading from the sequence of the most recent position fixes, this solution also works for stationary users. Magnetic compasses do not require extensive computational power, nor do they need any prior knowledge about the environment – both of which is the case for approaches based on optical image recognition [12],[18]. First mobile phone models with comparable embedded magnetic compass sensors are already being introduced by handset manufacturers (e.g. Nokia model 5140). The compass used in our prototype device can reliably resolve <0.07mGauss. Compared to the typical magnetic fields in the x- and y-horizontal plane, which are in the range of 200 mGauss (more at the equator, less near the poles), it achieves a theoretical resolution of 0.02°. This resolution is practically superposed by magnetic sensor errors, variations of the earth's magnetic field, nearby ferrous materials, A/D converter resolution errors and temperature effects (as discussed for instance in [5]), summing up to a total accuracy of about 1-3° and 0.1° resolution.

4.2 Tilt Sensor

The 2-axis tilt sensor detects the mobile phone's pitch and roll angles (compare figure 3) by measuring the acceleration that is exerted on a small mass (or heated gas) by the earth's gravitational field. The mechanic principle of the sensor makes it particularly robust against external influence, such as electrical interference or magnetic fields. Compared to gyroscopes or electrolytic fluid based sensors, no rotating or pendulous parts are required, allowing smaller structures and higher resonant frequencies. The operating frequencies of such sensors easily achieve several 100Hz, thereby allowing the detection of fast user movements.

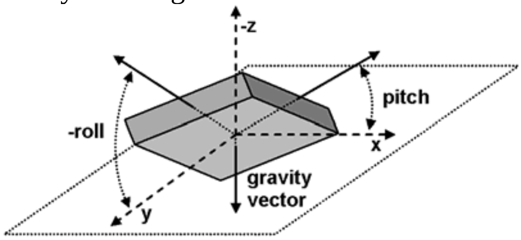


Fig. 3. Tilt angle definitions

Similar to the heading error, the tilt error of the sensor used in our prototype is less than 1°, resulting from temperature effects and measurement noise [1]. The acceleration sensor reaches maximum sensitivity when held normal to the gravitational field (parallel to the earth's surface). Since the tilt sensor used in our prototype is based on a single acceleration sensor for each axis, the resolution therefore decreases for tilt/roll angles above 60°. Consistent resolution of 1° for the entire range from 0 to 360° can be achieved by combining the measurements of 2 sensors arranged perpendicularly [1].

4.3 Related Work

Hinckley et al [8] demonstrated how tilt sensor data can be exploited locally on a PDA device. His experiments were aimed at proving that orientation detection can be used beneficially as a novel input modality that makes single-handed operation more natural and intuitive. Examples from his experiments include the activation of screen-scrolling by tilting the device or the activation of the PDA's voice recording functionality by holding it to the mouth and ear like a phone. Hinckley's experiments also included discussions and measurements of error probabilities and false-positive detection for certain gestures, such as the voice-recording activation gesture. Similarly, Eslambolchilar and Murray-Smith [6] implemented tilt-controlled zooming and scrolling on a PDA device. Their test users found that tilt-based single hand control was an intuitive solution to the problem of navigating large documents on small displays. We implemented a number of applications that also use the orientation of the device for single-handed input. One example is an experimental concept user interface with tilt-controlled slide-in menus. Another example is a dexterity game where the user must guide a ball through a maze without dropping it into holes in the maze floor. A photo of the device prototype, together with pictures of both applications is shown in figure 4.

Fig. 4. Device prototype and demonstrator applications

Due to an obvious potential for mobile entertainment and gaming (which has already turned into an important revenue channel for mobile network operators, with an estimated market worth $1.2 billion by the year 2006 [3]) and the possibilities for novel input modalities like gesture recognition, handset manufacturers have recently announced a number of mobile phone models which will be equipped with comparable tilt or acceleration sensors. Similar plug-in sensors for PocketPC PDAs are also available commercially.

5 Creative Histories Mobile Viewer

In addition to the local application of orientation awareness, as described in the previous section, we also aim to demonstrate the technical feasibility of the full Point-to-Discover concept in a practical setting: The objective of the "Creative Histories" project [2],[16] is to create a high-quality three-dimensional model of a cultural heritage site. The area being re-created is a square in downtown Vienna, Austria: the Josefsplatz. The virtual 3D model of the Josefsplatz is thereby not just confined to its current constructional state. Rather, it encompasses the constructional state in different historical stages throughout history: Users of the system will be able to navigate through the model in real-time and virtually move back and forth in time. A second objective of the project is to associate a virtual information space with the 3D model: Users can quickly and intuitively retrieve different types of historical information and media (such as textual information, historic images, photos or audio and video documents) related to certain locations, such as buildings or landmarks on the square.

Based on the device prototype introduced above, we are implementing a client application for a mobile smartphone device. With this application, the phone acts as a "real-world navigation tool" to the Creative Histories system: The user's field of view on the square is continuously determined by the on-device GPS and orientation sensors. By downloading appropriate chunks of 3D geometry and suitable textures from the geometry server, the mobile device can render a live simulation of what the user's view looked like in a different historical epoch – the screen acts as an interactive "window to the past". The application also indicates when there's additional information available for a certain building or object. The user can access the information by pointing the device towards the real-world object and selecting it in the 3D simulation [4].

Figure 5 shows a view of the Josefsplatz together with an emulator screenshot from an early version of the viewer application, which features a yet untextured environment model. As can be seen, the application implements the Point-to-Discover metaphor described earlier: Using positioning and orientation detection, the mobile phone functions as pointing device in the virtual 3D model of the real world. Furthermore, the application extends the basic principle by also representing the 3D model itself visually on the screen.

Fig. 5. Creative Histories mobile viewer application (untextured model)

6 Acknowledgements

Project Creative Histories is funded by the WWTF funding programme. It is a collaborative research project of the Austrian National Library, the

Austrian Research Institute for Artificial Intelligence (OFAI), the Research Center for Virtual Reality and Visualization, Ltd. (VrVis), and the Telecommunications Research Center Vienna (ftw.).

7 References

[1] "Accelerometer fundamentals", memsic application note, AN-00MX-001, http://www.memsic.com/
[2] "Creative Histories – The Josefsplatz Experience" project homepage. http://www.josefsplatz.info/
[3] "Sega, AT&T Wireless Push Mobile Gaming." Wireless News Factor Press release, http://www.wirelessnewsfactor.com/perl/story/19673.html
[4] Baillie, L., Kunczier, H., Anegg, H., "Rolling, Rotating and Imagining in a Virtual Mobile World." 7th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 05), Salzburg, Austria, September 19-22, 2005.
[5] Caruso, M., "Applications of Magnetic Sensors for Low Cost Compass Systems." Honeywell SSEC, http://www.magneticsensors.com/
[6] Eslambolchilar, P., Murray-Smith. R. "Tilt-Based Automatic Zooming and Scaling in Mobile Devices – A State-Space Implementation." Proceedings of the Mobile HCI conference (Mobile HCI'2004), Glasgow, Scotland, September, 13-16, 2004. Springer-Verlag, Heidelberg, 2004, 120-131.
[7] Frueh, C., Zakhor, A., "Constructing 3D City Models by Merging Ground-Based and Airborne Views." IEEE Computer Graphics and Applications, Vol. 23, no. 6, pp. 52-61, November/December 2003.
[8] Hinckley, K., Pierce, J., Sinclair, M., Horvitz, E. "Sensing Techniques for Mobile Interaction", 13th ACM Symposium on User Interface Software and Technology, San Diego, California, USA, 2000.
[9] Hu, J., You, S., Neumann, U., Park, K. K. "Building Modeling from LIDAR and Aerial Imagery." ASPRS'04, Denver, Colorado, USA. May 23-28, 2004.
[10] Kaasinen, E., "User Needs for Location-Aware Mobile Services." Personal and Ubiquitous Computing, Volume 7, Issue 1, May 2003.
[11] Karner, K., Bauer, J., Klaus, A., Schindler, K., "MetropoGIS: a city information system.", IEEE Conference on Image Processing (ICIP'02), Rochester, USA, 2002
[12] López de Ipiña, D., Mendonça, P. R. S., Hopper, A., "TRIP: A Low-Cost Vision-Based Location System for Ubiquitous Computing." Personal and Ubiquitous Computing, Vol. 6, pp 206-219, Issue 3, May, 2002.
[13] Nevatia, R., Price, K., "Automatic and Interactive Modeling of Buildings in Urban Environments from Aerial Images." IEEE Conference on Image Processing (ICIP'02), Rochester, USA, 2002.
[14] Parkinson, B. W., Spilker Jr., J. J. (ed.), "Global Positioning System: Theory and Applications", Vol. 2, American Institute of Aeronautics and

Astronautics, Inc., 370 L'Enfant Promenade, SW, Washington, DC 20024-2518, 1996

[15] Rao, B., Minakakis, L., "Evolution of Mobile Location-based Services." Communications of the ACM. December 2003. Vol. 46, No. 12.

[16] Sormann, M., Reitinger, B., Bauer, J., Klaus, A., Karner, K., "Fast and Detailed 3D Reconstruction of Cultural Heritage." International Workshop on Vision Techniques applied to the Rehabilition of City Centres, 2004.

[17] Syrjärinne, J., "Studies of Modern Techniques for Personal Positioning", Doctoral Thesis, Tampere University of Technology, March 2001.

[18] Ulrich, I., Nourbakhsh, I., "Appearance-Based Place Recognition for Topological Localization." IEEE International Conference on Robotics and Automation, San Francisco, USA, April, 2000, pp. 1023-1029.

[19] Wasinger, R., Stahl, C., Krüger, A., "M3I in a Pedestrian Navigation & Exploration System". 5th International Symposium on Human Computer Interaction with Mobile Devices (MobileHCI 2003), Udine, Italy, September 8-11, 2003.

Geo-Services and Computer Vision for Object Awareness in Mobile System Applications

Patrick Luley[1], Lucas Paletta[1], Alexander Almer[1], Mathias Schardt[2], Josef Ringert[3]

[1] Joanneum Research, Institute of Digital Image Processing, Graz, Austria
[2] TU Graz, Institute of Remote Sensing and Photogrammetry, Graz, Austria
[3] TU Graz, Institute of Geoinformation, Graz, Austria
patrick.luley@joanneum.at

Abstract

In recent years, location and context aware systems have been presented for the indexing and annotation of both location and user state relevant information to the user. These systems were mostly based on georeferencing from GPS signals, and driven by changes of the system state in dependence on situations that impact the overall system performance (user position, energy consumption, responsiveness, etc.). In this chapter we focus attention on a completely innovative aspect of contextual indexing for mobile system applications. We claim that the application of geo-services and vision based sensing enables systems to determine object characteristics information that can be used for semantic indexing of information and thereby provide an innovative quality of service.

Computer vision can be used to extract object information, such as the identity of buildings (Figure 1), information signs, people, etc., and thereby enable object based indexing which can finally be applied to extract a semantic description of the environment. Geo-services are mandatory to support vision based object recognition and semantic

indexing in several ways. Firstly, location based sensing provides information about the geo-reference of objects in the field of view, it reduces the number of possible object hypotheses and therefore can dramatically simplify the complexity of object recognition. Secondly, vision provides an estimate of the user position from the geo-referenced object information. Geo-services can use this to reason more efficiently about a current object and user position.

1 Introduction

Computer vision is an emerging technology for providing reliable vision competences outdoors, dealing with the immense complexity of the visual input and associated degrees of freedom in image understanding. Mobile imaging as an upcoming technology will become ubiquitous such as in camera phones, vision enhanced handhelds, and wearable cameras. Mobile vision technology would make new approaches possible in scenarios of personal assistance, mobile work, assistive services and also in general location based services.

Services, based on location awareness, require the knowledge about the actual location of the user, the current user context and geo-referenced information about areas and points of interest. Different technologies can be used to fulfill these requirements. Location awareness can be provided based on GPS, using wireless network technologies such as GSM and WLAN or using self-location possibilities, e.g. based on street names and house numbers. Location awareness for a mobile service in urban areas can, by the use of GPS only, not be assured everywhere and anytime, because of the known weaknesses of GPS signal availability in urban areas. Mobile systems operating in urban environments must take advantage of contexts arising from the spatial and situated information at a current location of the pedestrian user. Therefore, location awareness can be realized based on the knowledge of the location of the geo-referenced objects of interest, which allows also determination of the user position.

In frame of the projects MOBVIS (“Vision Technologies and Intelligent Maps for Mobile Attentive Interfaces in Urban Scenarios”, EC funded project) and “Mobile City Explorer”, which is a national funded project (Austrain Space Agency), a scenario of city exploration (see Fig. 1) is proposed [LPA05], where the tourist pedestrian can send visual queries to a server based object recognition system. The goal of the system is to attain situated object awareness, i.e., to get the semantic information about the environment for object and location identification, to get semantically indexed access to knowledge data bases. This methodology has the advantage that the object information is most appropriate to structure perception, and to structure the indexed search into annotation information.

In the following chapters a mobile application system (see Fig. 2) will be described which offers area-wide location awareness based on a common smart-phone and image based object recognition tool in combination with a GPS module.

2 User Scenario

This section briefly describes a user scenario, in case of a city-tourist type pedestrian, focusing on the service of image based object recognition.

The common way of doing city sightseeing is using a printed city map with integrated sightseeing-tours leading the tourist along a pre-defined path from one sight to the next. Brief descriptions of the sights can be found at the backside of the map.

By the use of the image based object recognition service the tourist gets the freedom to explore the city without any pre-defined sightseeing tours. The tourist moves completely free through an unknown area and if he is interested in any object (e.g. a historical building or a statue), he just has to take a picture of it with his camera-phone, with or without GPS device connected, and press the “Identify” button. As result he gets a detailed description of the object containing multi-media tourist information. As a second achievement he also gets the position of the identified object which can be used for navigation. As the position is well known from the GPS device or the object recognition the Geo Services are able to render a map showing the surrounding area. This map is also sent to the tourist as part of the response of the object recognition and gives him an overview of the near ambience.

Our user scenario and the role of image recognition in mobile tour guides was well-investigated in a usability study by Nigel Davies, Keith Cheverst, Alan Dix and Andre Hesse [DCDH05]. The result was that there is a very high acceptance for image based object recognition by tourists.

Fig. 1. Scenario: capturing an image about a building object of interest that is to be identified by the system.

3 System architecture

We will briefly describe how a common smart-phone with built-in digital camera can be used for image based object recognition. A GPS device, built-in or connected to the phone via cable or Bluetooth, can help to accelerate the recognition process considerably. The overall concept consists of three main phases.

In the first phase, a software client is activated by a user on his personal smart phone. The software can be directly downloaded from a website to the internet enabled smart-phone and will then be installed on the device. The software-client offers functionality to capture an image of an object the user intends to identify. Next, if available, the smart-phone reads the actual position from the GPS device. If the GPS cannot obtain a position for any reason, the cell information of the phone-network provider can be used to approximate the user location instead. The image of the object and the position of the snapshot-location are combined together into a SOAP Message and sent to an image-recognition web-service, running on a dedicated server, over a common wireless internet connection like GPRS or UMTS.

In the second phase, the web-service reads the request from the client (smart-phone) and extracts the image and the GPS position. The image is then analysed by a dedicated recognition algorithm to obtain representative image features like local descriptors that characterize a specific object. Next, these features are compared with reference features stored in an object-database. The database contains images of different objects, with the pre-processed features and the position of the snapshot-point. The object is then identified by matching the features of the user-picture with the features of the database. This matching process can become, in the case of a large database with many objects, very time consuming and that is where the GPS position can help. To accelerate this process the set of objects under investigation in the database can be selected with the user’s GPS position. Only those database objects come into account for the matching process, which have corresponding images with a snapshot-point near the user’s position. (Fig. 3) Once the object is identified, the object-database is queried for some information containing text and pictures. This information is integrated into a SOAP-message and is then sent to the smart-phone client as response.

In the third phase the web-service response is presented to the user on the smart-phone. Additional to the object quick information, containing text, pictures and a location based city-map, there is a URL which can be used by the user to obtain detailed information about the object. The URL can be viewed in a common smart-phone internet browser like “Opera” or “Pocket Internet Explorer“. The URL contains, as a parameter, the unique identification number of the desired object and links to a dynamic website, which is generated at runtime on the server. The layout of the website is optimized for the requesting hardware platform – because different smart-phones have different display resolutions. Fig. 2 illustrates the complete technical concept and its three phases.

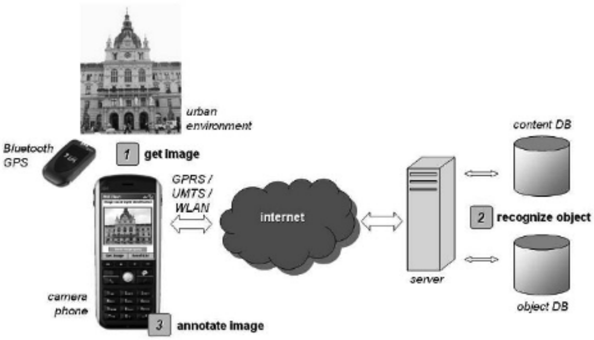


Fig. 2. Overview of the mobile application systems concept

The aim of this client-server architecture is to bring the image based object recognition service to any person using a common camera-phone and to gain scalability in reference of the number of objects in the database and complexity of the image-recognition algorithms.

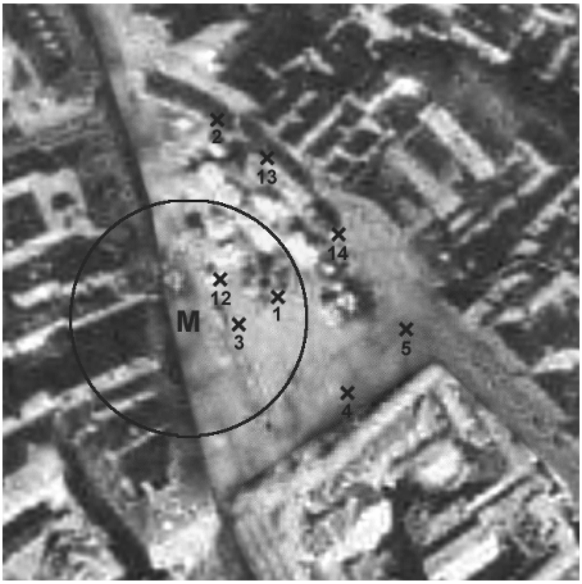


Fig. 3. Geo-Context, e.g., from GPS based position estimates ('M' with blue uncertainty radius), can set priors by geographically indexing into a number of object hypotheses ('X's are coordinates of user positions while capturing images about objects of interest).

4 Visual Object Recognition for Object Awareness

In contrast to signal based approaches as proposed in most location based services, object awareness points towards semantic indexing which will enable much more flexible interpretation of the information from a local environment. Object awareness relies on a structural matching process of comparing the situated information extracted from the sensor streams with prototypical patterns that were developed from experience. In the case of computer vision, object recognition methods enable to identify characteristic patterns of visual information in the field of view, such as infrastructure objects (tourist sights, traffic signs, information boards, etc.), people, or objects of every day's use (chair, lamp, mobile phone, etc.).

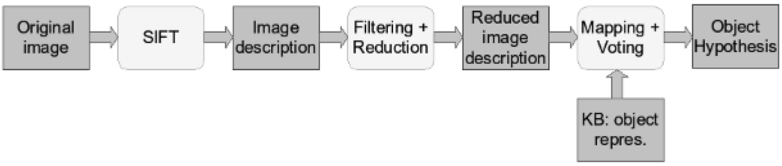


Fig. 4. Sketch of the methodology used for object recognition in a robust framework. Local patterns (SIFT) are detected for an image description, filtering using only informative descriptors leads to a reduced description. This sparse description is matched towards a reference description to determine object hypotheses.

Our proposed image based service for object awareness requires both, robust and fast visual object recognition of typically low-quality outdoor images. We therefore applied a methodology that is highly suited for mobile vision applications, i.e., the Informative Features approach [FSPB04]. This method extracts state-of-the-art local image features, i.e., SIFT (Scale Invariant Feature Transform) descriptors [LOW04], from the greyscale image, in a first step. These texture patterns are then analysed to identify the most informative descriptors among them which are retained to establish a kind of attention filter. Responses in terms of most informative features are then matched to reference descriptors to provide a probability distribution on a pre-determined set of object hypotheses. In addition, the Informative Features approach enables to detect any object presence in the image, and background vice versa [FSPB04]. We also provide a confidence value to quantify the degree of uncertainty in the final object decision, and thereby enable to define a 'quality of service' measure for usability analyses.

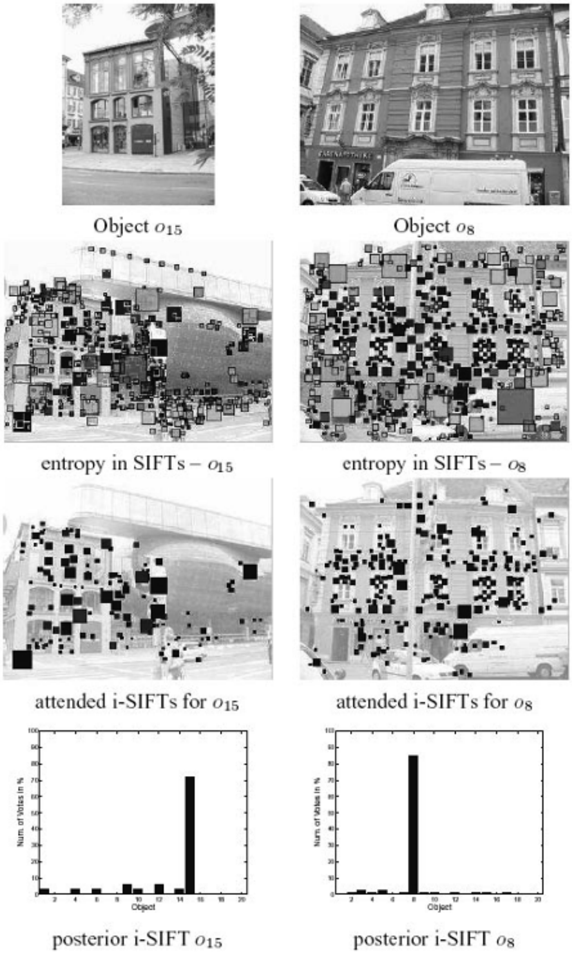


Fig. 5. Sample recognition of objects o_{15} and o_{18} ("Kunsthaus", "Bärenapotheke", out of the TSG-20 image reference database). Top down: (i) reference images, (ii) location of SIFT descriptors with colour coded quality of service, (iii) selection of high quality descriptors, (iv) distribution on object hypotheses with correct decision based on maximum confidence.

5 Geo Services

As already mentioned, the knowledge of the tourist position is a very important fact within the system, not only for routing and guidance purposes but also to support the matching algorithm for the object recognition. Primarily GPS is used for positioning. If GPS is not available, either the cell information of the phone-network provider, a successful accomplished object recognition, or a manual input of the street name and number can be used for position determination. In the later case the street name and number is matched against a database where all buildings are stored with their according coordinates. This is especially in urban areas a method where the user can determine his position with minimal effort.

The actual position is also used to present maps of the surrounding area of the tourist on the mobile device. The maps show a city map for orientation purposes, points of interest with continuative links, the GPS track of the device and routing information. For the creation a map server application running on the server is used, cf. Fig. 2. In the current projects the University of Minnesota Map Server (UMN) is used. The primary function of the map server is to read data from various sources, which not necessarily need to be stored on the same computer, and pull these layers together into a graphic file. One layer may be the city map, another the points of interest, or the routing information. Each layer is overlaid or drawn on top of the others and then printed into a graphic to be displayed on the mobile device. Due to the fact that the map content depends on the user profile and interests, a dynamic creation of the content is necessary. This is done by using PHP MapScript.

The map server application is able to easily integrate data stored in databases. For applications which bear strong relation to positions on earth the use of databases with support of geographic features is advisable. In a lot of commercial and open source databases this support can be found. In the context of the Mobile City Explorer project PostGIS is used. PostGIS adds support for geographic objects to the PostgreSQL object-relational database and follows the OpenGIS “Simple Feature Specification for SQL”. PostGIS enables the execution of spatial queries by offering spatial functions, such as Distance(), Intersect(), and Within(), among others, to narrow down the result of the search.

Cartography on mobile devices is not comparable with traditional or digital methods. The small display size, limited processor power of the device, and varying lightning conditions have to be taken into account. The map must show only the relevant degree of details for the current task. The dedicated use of colour and font size shall support the readability under different lightning conditions.

6 Conclusions

Object awareness provides a concept for semantic indexing into huge information spaces where standard approaches suffer from the high complexity in the search processing otherwise, and provide a means to relate the mobile agent to a semantic aspect of the environment. Visual object recognition using innovative and robust pattern recognition methodologies is an emerging technology to be applied in mobile computing services. Due to the many degrees of freedom in visual object recognition, it is highly mandatory to constrain object search by means of context based attention. In this paper we used geo-context to focus object recognition on a specific set of object hypotheses, and demonstrated that the challenging problem of identifying tourist sights in urban environments can become feasible.

In the projects MOBVIS and Mobile City Explorer our work will target at further exploiting contextual relations in order to cut down the complexity in otherwise unconstrained visual object recognition and make mobile services feasible. This challenging research goal will be achieved by identifying and tracking spatio-temporal aspects of the user's task context, improving the representation of geo-context by multi-sensor information fusion, and by applying attention strategies on the visual input. In addition, future research results will be achieved in more developed demonstrators, using camera equipped phones and wearable interfaces as mobile platforms for presentation.

Acknowledgements

This work is funded by the EU-IST project MOBVIS and the ASA via project "Mobile City Explorer" coordinated by Arsenal Research.

References

[LPA05] Luley, P., Paletta, L., Visual Object Detection from Mobile Phone Imagery for Context Awareness, Proc. mobileHCI2005, Salzburg, 2005, pp. 385-386.

[DCDH05] Davies, N., Cheverst, K., Dix, A., Hesse, A., Understanding the Role of Image Recognition in Mobile Tour Guides, Proc. mobileHCI2005, Salzburg, 2005, pp. 191-198.

[LOW04] Lowe, D., Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60(2), pp. 91-110, 2004.

[FSPB04] Fritz, G., Seifert, C., Paletta, L., and Bischof, H., Rapid Object Recognition from Discriminative Regions of Interest, Proc. 19th National Conference on Artificial Intelligence, AAAI, San Jose, 2004, pp. 444-449.

Section IV Visualisation and Cartographic Communication

Chapter 23	Verena Radoczky How to design a pedestrian navigation system for indoor and outdoor environments
Chapter 24	Ken Francis, Peter Williams Dancing_without_gravity: A story of interface design
Chapter 25	William Cartwright Landmarks and the perception of a space in web-delivered 3D-worlds
Chapter 26	Michael Wood, David G. Pearson, Colin Calder Comparing the effects of different 3D representations on human wayfinding
Chapter 27	Jürgen Döllner, Benjamin Hagedorn, Steffen Schmidt An Approach towards Semantics-Based Navigation in 3D City Models on Mobile Devices
Chapter 28	Dieter Schmalstieg, Gerhard Reitmayr The World as a User Interface: Augmented Reality for Ubiquitous Computing
Chapter 29	Tobias Höllerer, Jason Wither, Stephen DiVerdi “Anywhere Augmentation”: Towards Mobile Augmented Reality in Unprepared Environments
Chapter 30	Rex Cammack Open Content Web Mapping Service: A Really Simple Syndication (RSS) Approach

How to design a pedestrian navigation system for indoor and outdoor environments

Verena Radoczky

Research Group Cartography, Department of Geoinformation and Cartography, Vienna University of Technology
radoczky@surfeu.at

Abstract

The fast transfer rates via UMTS and their corresponding high-tech devices provide the opportunity to develop easily accessible Location Based Services, which can be utilised to guide pedestrians in unfamiliar environments. With the help of cartographic multimedia data navigation systems can be responsive to the needs of pedestrians by presenting information in the preferred forms. In order to facilitate ongoing tasks, not only the presentation forms themselves are decisive, but also their combination and deliberate design is a crucial factor for usability. Moreover landmarks are an important enhancement for pedestrian navigation systems, yet their immense diversity makes it difficult to effectively include them in guiding instructions. For that reason this paper discusses the use of active and passive landmarks, as well as multimedia presentation forms and general design goals for a combined indoor/outdoor pedestrian navigation service.

1 Introduction

Within the last few years car navigation systems have started to conquer the market in the Western World. Car drivers have started to trust in the guiding instructions of the system, and once they got used to it, they do not want to miss it anymore. Other than that navigation systems are not very popular until today. Pedestrians usually consult paper maps when looking for a destination. Some prototypes of portable tourist information systems (Modsching et al. 2005, Uhlirz & Lechthaler 2002, Zipf et al. 1999) are available today, but user acceptance does not seem to be very high so far (Forum Mobilkommunikation 2002, Kölmel & Wirsing 2002). Beside the costs possible reasons for this might be the lack of information provided to the user. In contrast to car drivers, pedestrians do not need to be as concentrated on traffic and can therefore process a lot more information. Often a pedestrian does have a little bit of extra time and would appreciate some additional information about his current environment, which is usually not included in today's commercial navigation software. This type of auxiliary information could be presented best in the form of multimedia-based data, which has hardly been taken into account so far. It should also be considered that the moment special stops at shops or sights are included, operating the program could become very complicated and demands a lot of patience from the user, which is why an emphasis on user interface design needs to be regarded.

Moreover the range of possible paths for pedestrians is a lot higher than the street graph based routes of vehicle drivers and it gets nearly impossible to create a network that includes them all, especially when expanded with indoor paths. Indoor navigation as such is yet a very new and unexplored branch of navigation. Some museums use PDA-based navigation systems to guide visitors through an exhibition (Oppermann 2003, Chan et al. 2005) but they merely concentrate on this very specialised application area. Also wayfinding experiments within airports have been conducted by Raubal (2001) which give a deeper insight into preferred paths of travellers and route choices at public transport stations, but research on visualisation aspects of combined indoor and outdoor routes is rare (Baus et al. 2002). Contrary to outdoor urban environments, where the third dimension can mostly be neglected when visualising a route, buildings require the depiction of floor levels, staircases and elevators. This demands either the switch to a 3D presentation form or a well considered design of floor plans. Since it is more likely to lose orientation within a building than outside (Hohenschuh 2004, Radoczky 2003), it is also desirable to use indoor landmarks which on their part are harder to define than outdoor landmarks.

Nevertheless route guidance within large buildings like shopping centers, universities or official buildings in junction with outdoor routes would often be desirable. Users often get into the situation where they arrive at a subway station and look for the way to a specific room in a nearby public building. With the help of multimedia presentation forms this task could be assisted by a user-friendly navigation system.

Additionally to navigation support it could be beneficial to supply the user with information that is adapted to the current task. For instance when visiting a shopping centre, information about bargains at favoured shops could be displayed, or when strolling around an airport or train station, information about departing planes or trains that concern the user could be provided. The technical requirements for these types of online services are mostly not achieved today, but it is imaginable that future systems will provide them in more variety.

2 Preparation and applicability of various multimedia presentation forms

Many different communication forms are imaginable when supporting pedestrian wayfinding in urban environments. Each of them has advantages and disadvantages that need to be taken into account when combining them.

2.1 Maps

Tests have shown that maps are the most important presentation form when communicating spatial information (Reichl 2003, Radoczky 2003). Even though maps can differ dramatically from the perceived structure of a spatial environment, they can help the user to get a good overview of the area. Precision does not seem to play an important role in the navigation process, as long as essential topological information can be extracted and distortion is not disproportionately high (Dennett 1969, Agrawala and Stolte 2000, Evans 1980, Carstensen 1991, Moar and Bower 1983, Klippel 2003). Yet the design of a map can differ dramatically depending on the region where they come from. City maps in the USA typically show a network of street graphs whereas European maps usually depict building blocks. Sometimes users prefer a more sketchy style (Döllner & Walther 2003) and users who are very familiar with an area could even favour a schematic map that concentrates on topology rather than on topography (Agrawala & Stolte 2000, Radoczky 2005). Therefore the design of a map needs to be adapted to the region and the purpose of use.

Moreover the map should be available in different scales with different levels of detail. That way the whole route which is demanded by the user can be shown on the display of the device as an overview map without any need to scroll or pan. The moment guidance begins, an automatic zoom can switch to a larger scale with more detail, maybe even including house numbers. During navigation it would also be advisable to automatically twist the map to an egocentric map view, in which case the north direction should always be marked on the map (Radoczky 2003, Zipf & Jöst 2004, Hohenschuh 2004). That way mistakes at crossings can be avoided because turning points can be viewed as simple left and right turns. Certainly it would be possible to twist and turn the usually small handheld device by the user himself, but then labels and other additional information would lack readability.

Another explicit advantage of maps in comparison to pure textual or oral guiding instructions is the possibility to include multimedia features with the help of interactivity. Hot spots can act as obvious hiding places for landmark information and other additional information sources (Uhlirz 2001, Nagi 2004).

2.2 Floor plans

The map-like depiction of different building levels as an adapted floor plan seems to be a good possibility to represent complex indoor structures and so far they have also proven to be the most effective ones (Radoczky 2003, Ortag 2005, Werner & Long 2003). Even though they have similar properties as maps (and are therefore rather easy to read), generalisation techniques need to be customised to the specific nature of indoor environments. Since indoor navigation is a relatively new sector in LBS (Location Based Services), there are no unified regulations for design modes of floor plans. Thus legend symbols and colour management are usually arbitrarily chosen or adjusted to the symbolisation of CAD-programs. It is therefore important to adapt symbolisation to signage of the respective country in order to increase the recall value within the building. This aspect becomes especially important when representing the switch between different floor levels while walking up/down stairs or using an elevator. To overcome the problem of vertical movement while using a 2D map, universally understandable pictograms that symbolise stairs, elevators or escalators can be displayed (see figures 1-3).

Overall the generation of floor plans demands a similar expenditure of time as the production of a map. Concerning the graphic type Butz et al. (2001) recommend the usage of vector graphics because they have the advantage to facilitate incremental transmission to a mobile device. This method is especially useful when transmission rates are rather low. The most essential elements can then be transmitted first while details of the environment appear little by little.

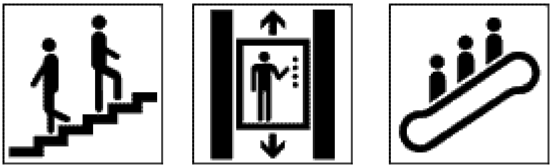


Fig. 1. Symbols for stairs, elevators and escalators

2.3 Verbal guidance

Verbal guiding instructions have also shown to be a helpful tool for pedestrians. A test by Ortag (2005) indicated that test candidates who had to find part of a route with the help of a map and the other part with audio guidance did not show major performance changes after switching the presentation form. Only when being asked about the preferred mode, a slight tendency towards map presentation was observable.

When being used to a personal device with earphones, oral information can be very useful in quiet environments because noise disturbance can be kept to a minimum. In noisy areas it can be difficult for the user to understand the electronic voice. Sentences should then be repeated until the information is definitely perceived by the user, which decreases the effectiveness of this method.

Concerning the configuration of “good” route instructions, some guidelines should be followed (Lovelace et al. 1999, Habel 1987, Kray et al. 2003):

- Users should be prepared for upcoming decision points.
- Instructions should be kept simple and clear.
- Landmarks should be included in descriptions.
- Street names (or corridor numbers) should be mentioned, but only, if they are usually displayed in the network of the respective city (or building).
- In case information about the segment is rare, distances can be mentioned.
- The user should have the possibility to replay the last sentences.

The realisation of an audio guidance system is rather simple. Different programs convert text into voices or in case the amount of words extends a certain limit, concatentive synthesis can be applied, where real voices are divided into individual parts and can then be assembled in another sequence (Lemmetty 1999).

Textual guidance is the most simple presentation form for navigation systems. They are easy to create and do not require a lot of memory. Guidelines for their configuration are very similar to audio guidance, with the advantage that sentences must not be kept to a minimum and can be written out in full and even when the user walks along a busy street (or some other noisy environment) he will be able to read the text (Reichl, 2003). On the other hand, darkness and weather conditions can make it impossible to read written text on a small screen and another main disadvantage of written text is the problem that the user spends more time reading and might not be as attentive to his environment.

Recapitulating we can say that both, written and audio instructions, do not demand a lot of technical resources and are therefore easy to implement. Yet in comparison to visual presentations their main disadvantage is, that they purely concentrate on the communication of the route and do not convey an overview of the whole area.

2.4 Images

An image can be a helpful presentation form in a navigation system, even though it is not really valuable as a navigation aid (Radoczky 2003, Heidmann & Hermann 2003). Users need a lot of time to compare reality with the photograph which is why it is advisable to merely use them when describing start point, destination or landmarks. Even here it should be ensured, that the depicted object is very distinctive. Furthermore it should be clarified, that the area does not underlie any seasonal changes before including it in a guiding system. Parks usually have another appearance in winter than in summer time, markets are sometimes only open at certain months and Christmas decorations could also extremely change the appearance of a place.

To be on the safe side, photographs could be an optional choice that people, who are not familiar with the environment and who are not in a hurry, can choose to gain additional information. The same holds for panorama views. They contain a lot more information than ordinary photographs and they can look very impressive, but since not many users are familiar with panorama views (Radoczky 2003), they should only be scarcely used to avoid confusion.

2.5 Videos

Videos have similar properties as photographs and their potential as route information aids can be rated as rather low. Objects and streets can be directly compared and identified with reality, but the quick movement of the film often provokes the loss of orientation because it does not give a lot of time to watch everything in detail (Radoczky, 2003). In a survey by Reichl (2003) subjects noted that they would not appreciate a navigation system that uses videos. Nevertheless videos can be used as optional information about landmarks and sights. Historic movies may give an insight on what a place looked like in former days, videos of special events that took place at the referred spot, animations or time lapses could inform a tourist about the area. When preparing such video information for a mobile system, it is very important to keep in mind, that movies should only show the most important aspects to satisfy the user's interest and not lead to boredom. Moreover they should be kept to a minimum length and size, and data compression should be very high to guarantee full transmission of the whole movie within a short time, even if transmission rates are rather low.

2.6 3D presentation

Generally 3D presentation forms are very effective representations of space. They can give a good overview of our environment and can contain a lot more detail than traditional maps. Especially the visualisation of buildings could be made more effective and unambiguous than floor plans. Unfortunately it is very difficult and sometimes even impossible to use them as navigation aids. One of the main disadvantages lies in the large data files and their high requirements on memory space, but since memory rates are developing very fast, it is assumed that this problem can be overcome in the near future. A more precarious problem is the demand on display size that is necessary to present the complexity of a three-dimensional file. Today users are used to mobile phones the size of a credit card and definitely do not want to carry around a device larger than a standard PDA. It is though possible, that future devices will look different, like electronic paper or glasses with installed augmented reality systems. In that case the usage of 3D visualisation will have to be reconsidered.

2.7 Online services

Even though their availability is scarce today, online services that provide the user with additional up-to-date location based information are desirable. Similar to landmarks, information could be accessed via a simple mouse click on the display where an information window pops up to inform the user about special menus of restaurants, upcoming shows in cinemas, availability and prices of tickets or opening hours of shops. This form of online advertisement could become very profitable to businesses the higher the distribution rate of location based services rises. Moreover locations of pharmacies that are open at night and timetables of public transport lines could be of use for pedestrians. Equivalent to traffic information in car navigation systems, pedestrian navigation systems could provide similar services like business disruptions of public transport or construction sites at diverse sights.

As can be seen from the descriptions above, the basis of a pedestrian navigation system could either be a map (or respectively a floor plan for indoor environments), a pure audio or a pure textual guiding system, in which case additional media is negligible.

In case maps are used to visualise the surroundings, it is though possible and even desirable to further assist this presentation form with verbal guidance. Singular information can cause irritation, but too much information could also slow down decision making (Klippel, 2003), which is why either textual or audio instructions should be added as secondary guiding aids. Furthermore distinctive images at difficult decision points could pop up on the display of the user. With the help of an interactive map additional photographs, videos or online services could be accessible via a mouse click. That way they do not interfere or distract from the navigation task and can be viewed whenever the user demands additional information about his environment.

3 Integration of active and/or passive landmarks

In wayfinding tasks people tend to structure space with the help of landmarks. Whenever we are asked for the way, we naturally use these significant objects to ensure that our descriptions are well-defined. It is thus not surprising, that the importance of landmarks for wayfinding has extensively been discussed in several publications (e.g.: Michon & Denis 2001, Foltz 1998, Lynch 1960, Golledge 1999). Yet the immense diversity of landmarks makes it difficult to include them in navigation systems. Furthermore the determination of landmarks is a subjective process that differs from one person to another (Gartner et al. 2005) and therefore the identification of landmarks in the actual environment by an individual cannot be guaranteed. Some outstanding buildings, though, can act as universal landmarks that are distinctive to almost everybody and thus should be included as orientation points in the guiding instructions. Their visualisation should be adapted to their function. A shop or a restaurant is represented best by its individual label, whereas historic buildings could be visualised with the help of pictograms.

Since it is more likely to lose orientation within a building than outside (Hohenschuh 2004, Radoczky 2003) and changes in direction happen more frequently, it is possible that the user demands a higher density of landmarks indoors than outdoors. Unfortunately it is expected that buildings dispose of a smaller choice of landmark categories. Potential candidates for indoor landmarks could be elevators, escalators, stairs, doors, plants, information boards, signs, fire extinguishers, etc. Even though they might not be highly remarkable they could still be essential aids for way descriptions.

The user should also be able to define his own landmarks by placing labelled flag-like symbols on the map. Moreover the system itself could remember start or destination points of elapsed routes and place the flags automatically. The best way to access additional information about the landmark is by using “on mouse-click” functions. A pop-up window could contain textual information about the object, a photograph or even a video. Accessible audio files could contain spoken text, significant sounds or music that helps to describe and understand the landmark. An opera house, for example, could be associated with arias and a church could be represented by the sound of its bell. In case verbal information is provided, it is important to keep it short in order to be able to follow and comprehend coherences.

Nevertheless the user of a navigation system needs some additional help to ensure he stays on the right track. A new approach to this problem is the usage of active landmarks. Unlike conventional landmarks they automatically build up a spontaneous radio connection the moment the mobile device reaches their range and identify themselves without any intervention by the user. Either the identification works via a unique ID or with the help of coordinates. Additionally to the supply of mere positioning data, the major advantage of active landmarks lies in the possibility to provide system-independent additional information about their environment. Ideally not only textual data, but also graphic or multimedia data can be transmitted to further assist the users’ orientation task and possibly even help to get to know the environment. Possible communication techniques for active landmarks are WLAN, RFID, IrDA and Bluetooth, of which each has its own advantages and disadvantages concerning range, data transfer rates, power supply and cost and therefore need to be adapted to the specific situation (Thienelt & Radoczky 2006). Especially in indoor environments active landmarks might be a reasonable method to supplement the guiding system with positioning and landmark information. Again this data could be accessible via a mouse-click function. The visualisation of the active landmark itself does not need to be different from conventional landmarks. Only in case it does not fulfil traditional landmark criteria, in terms of distinctiveness, a striking symbol could appear to signalise the possibility to access additional information.

4 Additional Design Goals

Including cartographic multimedia data in pedestrian navigation is a major step to increase user-friendliness of a navigation system. Yet considerations about how to visualise the route itself and on how to additionally support the usability of the system have to be made. Hohenschuh (2004) listed view principles that should be followed when designing a navigation system:

- Specially support in situations where nature does not provide orientation points or hints.
- Support both wayfinding styles, the independent and the guided style.
- Planning of route should be adopted to the respective circumstances.
- The system must function even if positioning signals are not always available.
- Public transport should be included in pedestrian navigation systems.

These are only a few aspects of route communication that should be taken into consideration. In the following design recommendations for a guiding system in an urban environment are asserted:

When starting the navigation application, the first thing a user expects is a visualisation of his current position on a map and an immediate request to disclose the desired destination without any further menu navigation (see figure 4). Style and colour of the You-Are-Here-marker should be self-explanatory and catchy to assure that the user is at any time in no doubt of his whereabouts on the map. The notification of a destination point could be realised in different ways. Speech recognition seems to be an easy and quick method to interact with the system, but surveys showed that subjects had bad experiences with interacting with their mobile phones by natural language and that they would prefer to type in a location with the hard keys of their device like they are used to from writing e-mails or SMSs (Zipf & Jöst 2004, Radoczky 2003). Beside the possibility to type in the full address of the destination, also points of interest like restaurants, public buildings or even the next cash machine should be recognisable by the system. The moment the target point is submitted, different route calculation algorithms should be started to offer a choice of diverse routes, such as the shortest, quickest or most scenic path that should also include public transportation. Figure 5 shows an example, where the most scenic route, the easiest path and a route that includes a public underground line are offered. Since the destination point is not known as a universal landmark (like in this case the St. Stephens Cathedral in Vienna which is visualised with the help of a pictogram), it is signalised with the help of a flag. Since all of the routes should be completely visible on the screen of the device in order to get an overview of the area and also a larger scale for navigating is desirable, different map scales should be recallable. Kray et al. (2003) believe that "[…] it is necessary to find a proper zoom factor for the map, which provides sufficient details, but does not overload the user with information. […] it is also necessary to find a proper level of detail for the annotations of the map, e.g. street names or the highlighting of certain buildings or areas." (Kray et al. 2003, p. 119). It is therefore useful to provide different levels of detail of each map that is included in the navigation system in order to provide reasonable zooming functionalities. The possibility to change the scale by the user himself must not necessarily be given, but could also be started automatically by the system in situations where a change of scale is reasonable according to the situation and the underlying data. To avoid loss of the survey perspective while zooming into detail, a dynamic zoom, where the change between detail and overview happens fluently and quickly, could be implemented (nDimension 2004).

After the user has decided which route to take, guiding can start immediately. Preferably the trail should be visible at any time, and the distinction between the past and the future path should be unambiguous. One possible solution to this visualisation problem could be to make the trail such that the past trail could be dyed in a very light colour. In case the user's position cannot be updated dynamically, it is possible to visualise the estimated position of the user with the help of the You-Are-Here sign along the present edge and only colour past edges in a transparent style (see figure 6). Moreover it is possible to avoid, that the current position of the user on the map moves out of focus, by regularly centring the user on the display with the help of an automatic scrolling functionality.

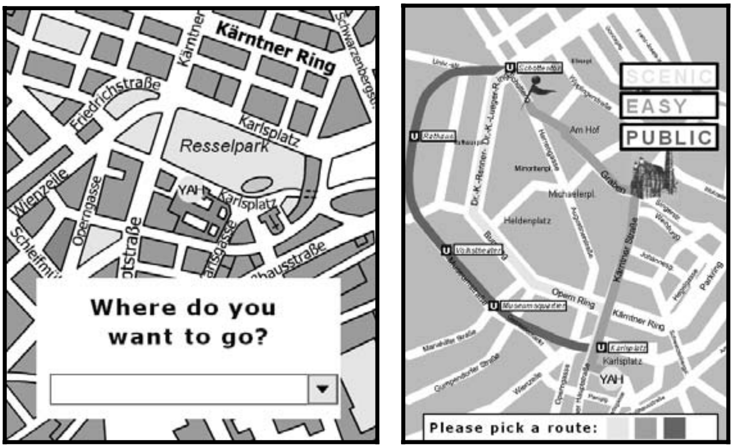


Fig. 2. Start of the application and choice of different routes

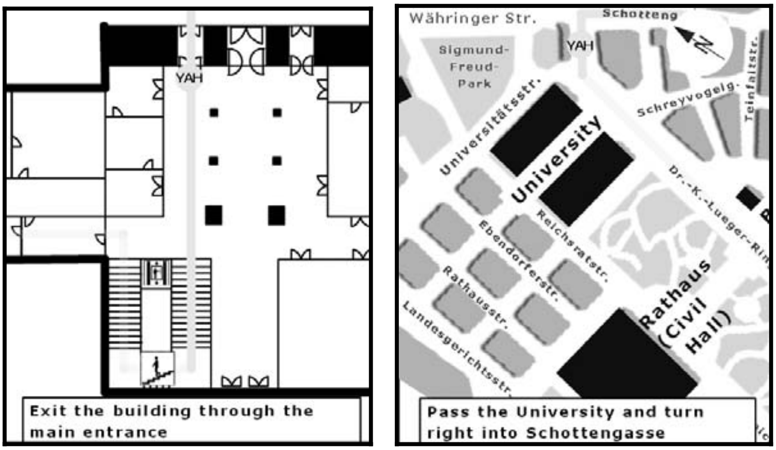


Fig. 3. Typical indoor situation and typical outdoor situation

In a combined indoor and outdoor navigation system, problems occur when the user enters or steps out of a building. The swap of the different presentation forms, the floor plan and the map, should look as seamless as possible to avoid reorientation problems on the user side. Again dynamic zooming has shown to be a highly effective method to overcome this problem. Figures 8-11 show some screenshots of a dynamic zoom while exiting a building:

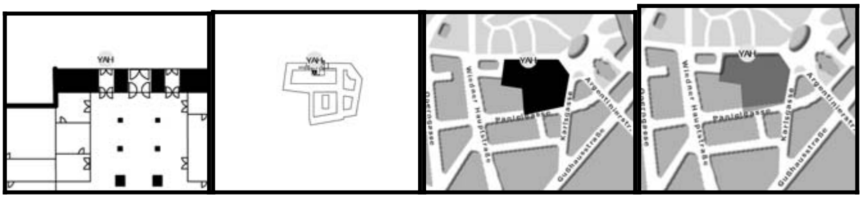


Fig. 4. Dynamic zoom while exiting a building.

While using an egocentric map view, some users prefer a permanent indication of the north direction, because it could help to keep orientation within the city. Also it is important to make sure, that labels and street names dynamically change to a readable position whenever the map was rotated automatically (see figure 7). This also applies for the marking of clickable objects, which on their side should consist of a striking colour that can easily be silhouetted from the background map.

5 Conclusions

Navigation systems that aim on pedestrians are scarce today. In many cases car navigation systems provide the possibility to switch to a pedestrian mode, but usually this mode is not as frequently used as the car navigation service. Preconditions of pedestrians differ enormously to those of drivers and the usage of the same data does not seem to suffice their requirements. Walking speed is a lot slower than driving speed and attention to traffic is not as important which gives the pedestrian a lot of time to concentrate on his environment. Because of this added possibility to process information, the potential to add multimedia presentation forms to the system that help the user to access more details and background information about his environment, can be utilised.

This contribution indicates, which presentation forms can be valuable for pedestrians and shows design goals to enhance user acceptance. Beside a map, where not only the route to be followed, but also landmarks and linked hot spots should be marked, verbal and textual guiding instructions and additional optional information in the form of photographs, panorama views or videos can be represented in a multimedia system. Also links to diverse online services are eligible that provide the user with up-to-date information about important objects and refer to special incidents and events.

Furthermore the integration of landmarks is discussed. The immense diversity sometimes makes it difficult to include landmarks in navigation systems, which is why also the usage of active landmarks, which automatically build up a spontaneous radio connection the moment the mobile device reaches their range, is considered. Additionally to the supply of mere positioning data, the major advantage of Active Landmarks lies in the possibility to provide multimedia information about their environment. Ideally not only textual data, but also graphic or multimedia data can be transmitted to further assist the users' orientation task and possibly even help to get to know the environment.

References

Agrawala, M., Stolte, C. (2000): A Design and Implementation for Effective Computer-Generated Route Maps. AAAI Symposium on Smart Graphics, March 2000.

Baus, J., Krüger, A., Wahlster, W. (2002): A Resource-Adaptive Mobile Navigation System. In: Proceedings of the 7th International Conference on Intelligent User Interfaces IUI02, San Francisco, January 13-16, 2002. New York, ACM Press, 15-22.

Butz, A., Baus, J., Krüger, A., Lohse, M. (2001): A Hybrid Indoor Navigation System. In: Proceedings of the 6th International Conference on Intelligent User Interfaces IUI01, Santa Fe (New Mexico, USA), January 14-17, 2001. New York (N.Y., USA): ACM Press. S. 25-33.

Carstensen K.-U. (1991): Aspekte der Generierung von Wegbeschreibungen. IWBS Report 190, Wissenschaftliches Zentrum, Institut für Wissensbasierte Systeme, 1991.

Chan, L.-W., Hsu, Y.-Y., Hung, Y.-P., Hsu, J. Y.-J. (2005): Orientation-Aware Handhelds for Panorama-Based Museum Guiding System. In: Smart Environments and Their Applications to Cultural Heritage, UbiComp '05, September 11-14, 2005, Tokyo, Japan.

Dennett D. C. (1969): The nature of images and the introspective trap. In: D. C. Dennett, Content and Consciousness, London: Routledge & Kegan Paul, pp. 132-141

Döllner, J., Walther, M. (2003): Real-Time Expressive Rendering of City Models. Seventh International Conference on Information Visualization, Proceedings IEEE 2003 Information Visualization, London, July 2003, 245-250.

Evans G. W. (1980): Environmental cognition. In: Psychological Bulletin, 88, 259-287, 1980.

Foltz, M. A. (1998): Designing Navigable Information Spaces. Master Thesis of Science at the Massachusetts Institute of Technology, June 1998.

Forum Mobilkommunikation, FMK (2002): Location Based Services: 5.10 Näher, mein Kunde, zu Dir…. Technische Grundlagen der Mobilkommunikation, Weißbuch Mobilkommunikation, last updated: October 6th, 2002: http://www.fmk.at/mobilkom/detail.cfm?Textid=70&Kapitelnr=8.

Gartner, G., Oakley, K., Jobst, M. (2005): Emotional Landmarks: An alternative view of way finding. Internal paper, publication in progress.

Golledge, R. (1999): Human Wayfinding and Cognitive Maps. In: Golledge, R. (ed.): Wayfinding Behavior, The John Hopkins University Press, Baltimore and London, 5 – 46.

Habel, C. (1987): Prozedurale Aspekte der Wegplanung und Wegbeschreibung. In: Schnelle, H., Rickheit, G. (eds.): Kognitive und neuronale Sprachverarbeitung. Westdeutscher Verlag, Wiesbaden.

Heidmann, F., Hermann, F. (2003): Benutzerzentrierte Visualisierung raumbezogener Informationen für ultraportable mobile Systeme. In: Contributios at the seminar „Visualisierung und Erschließung von Geodaten“, Hannover, February 2003. Deutsche Gesellschaft für Kartographie, Kartographische Schriften, Band 7. Bonn: Kirschbaum Verlag, 121-131.

Hohenschuh, F. (2004): Prototyping eines mobilen Navigationssystems für die Stadt Hamburg. Diploma Thesis, Faculty of Informatics, University of Hamburg.

Klippel, A. (2003): Wayfinding Choremes: Conceptualizing Wayfinding and Route Direction Elements; Doctoral Dissertation, Department of Mathematics and Informatics, University of Bremen.

Kölmel, B., Wirsing, M. (2002): Nutzererwartungen an Location Based Services – Ergebnisse einer empirischen Analyse. In: Zipf, A., Strobl, J. (eds.): Geoinformation mobil. Heidelberg: Herbert Wichmann Verlag, 85-97.

Kray, C., Elting, C., Laakso, K., Coors, V. (2003): Presenting Route Instructions on Mobile Devices. In: Proceedings of the International Conferences on Intelligent User Interfaces, January 12-15, 2003, Miami, FL, USA, 117 – 124.

Lemmetty, S. (1999): Review of Speech Synthesis Technology. Master Thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology.

Lovelace, K.L., Hegarty, M., Montello, D.R. (1999): Elements of Good Route Directions in Familiar and Unfamiliar Environments. In: Freska, C., Mark, D.M. (eds.): Spatial information theory: Cognitive and computational foundations of geographic information sciences, Proceedings of COSIT, August 25-29. Berlin: Springer-Verlag, 65-82.

Lynch, K. (1960): The image of the city. Cambridge, MA: MIT Press.

Michon, P.-E., Denis, M. (2001): When and why are visual landmarks used in giving directions? In: Montello, D.R. (ed.): Spatial Information Theory, Lecture Notes in Computer Science, vol. 2205. Springer, Berlin, 243-259.

Moar I., Bower, G. H. (1983): Inconsistency in Spatial Knowledge. In: Memory and Cognition, 11, 107-113.

Modsching, M., Kramer, R., Ten Hagen, K. (2005): Individuelle Stadtführungen mit einem Dynamischen Tour Guide (DTG). In: Strobl, J., Blaschke, T., Griesebner, G. (eds.): Angewandte Geoinformatik 2005, Beiträge zum 17. AGIT-Symposium Salzburg. Wichmann Verlag, Heidelberg, 457-466.

Nagi, R.S. (2004): Cartographic visualization for mobile applications. Thesis at the International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

nDimension (2004): http://nd.loopback.org/hyperd/visual.html, visited March, 29, 2004.
Oppermann, R. (2003): Ein Nomadischer Museumsführer aus Sicht der Benutzer. In: Szwillus, G., Ziegler, J. (eds.): Mensch & Computer 2003: Interaktion in Bewegung. Stuttgart: B. G. Teubner, 2003, 31-42.
Ortag, F. (2005): Sprachausgabe vs. Kartendarstellung in der Fußgängernavigation. Diplomarbeit am Institut für Geoinformation und Kartographie, TU-Wien.
Radoczky, V. (2003): Kartographische Unterstützungsmöglichkeiten zur Routenbeschreibung von Fußgängernavigationssystemen im In- und Outdoorbereich. Diploma Thesis, Department of Cartography and GeoMediatechniques, Vienna University of Technology.
Radoczky, V. (2005): Karten für Fußgängernavigation – kann die Bildung des mentalen Modells beeinflusst werden? In: Strobl, Blaschke, Griesebner (eds.): Angewandte Geoinformatik 2005. Beiträge zum 17. AGIT-Symposium Salzburg, 568-573.
Raubal, M. (2001): Human Wayfinding in unfamiliar buildings: A simulation with a cognizing design. In: Cognitive Processing (2-3): 363-388.
Reichl, B. (2003): Potential verschiedener Präsentationsformen für die Vermittlung von Routeninformation in Fußgängernavigationssystemen (FNS). Diploma Thesis, Department of Cartography and GeoMediatechniques, Vienna University of Technology.
Thienelt, M., Radoczky, V. (2006): Gegenüberstellung verschiedener Technologien für die Realisierung von Aktiven Landmarken. Internal paper, publication in progress.
Uhlirz, S. (2001): Cartographic Concepts for UMTS-Location Based Services. In: Proceedings of the 3rd Workshop on Mobile Mapping Technology. Cairo, Egypt.
Uhlirz, S., Lechthaler, M. (2002): LoL@ - City Guide. Prototyp einer kartenbasierten UMTS-Applikation. In: Geowissenschaftliche Mitteilungen, Schriftenreihe der Studienrichtung Vermessungswesen und Geoinformation TU Wien, Heft Nr. 58, 171-182.
Werner, S., Long, P. (2003): Cognition meets Le Corbusier – Cognitive principles of architectural design. In: Spatial Cognition 2003, 112-126.
Zipf, A., Häußler, J., Chandrasekhara, V., Malaka, R.(1999): GIS hilft Touristen bei der Navigation - ein erster Prototyp des mobilen Deep Map Systems für das Heidelberger Schloß. In: HGG-Journal (Journal der Heidelberger Geographischen Gesellschaft) , Heft 14, 285-296.
Zipf, A., Jöst, M. (2004): User Expectations and Preferences regarding Location-based Services – Results of a Survey. 2nd Symposium on LBS and TeleCartography, Vienna 2004.

Dancing_without_gravity: A story of interface design

Ken Francis, Peter Williams

Mapping Services Branch, Natural Resources Canada, Ottawa, Canada
kfrancis@nrcan.gc.ca

Abstract

An interface must perform just one function: to serve as the mediator between the user's wishes and the abilities of the machine. Good interface design has this goal as its beginning, but moves forward to advance the properties of ease, logic, elegance, consistency, wit and beauty in its form. The digital cartographic commodity as ideal map vernacular proposes an extension of the possibility of the paper map. This is only realizable if we remove the map from allegory and free it from metaphor. Interface metaphors have become redundant and clumsy, dampening the experience of "using the map". As much as possible, the map must serve as the means by which the user explores the information. It is through this confrontation with the programmatic agenda that a rich form/content dialectic can be engendered. These are the understandings that served as the starting point for problematizing the design of interfaces used in decision-making environments developed in the Earth Sciences Sector of Natural Resources Canada and the Canadian Geomatics industry. This chapter will discuss these approaches.

1 Introduction

An interface must perform just one function: to serve as the mediator between the user's wishes and the abilities of the machine. Good interface design has this goal as its beginning, but moves forward to advance the properties of ease, logic, elegance, wit and beauty in its form.

The digital cartographic commodity as ideal map vernacular proposes an extension of the possibility of the paper map. This is only realizable if we remove the map from allegory and free it from metaphor. Interface metaphors have become redundant and clumsy, dampening the experience of "using the map". As much as possible, the map must serve as the means by which the user explores the information. It is through this confrontation with the programmatic agenda that a rich form/content dialectic can be engendered.

These are the understandings that served as the starting point for problematizing the design of interfaces used in a decision-making environment developed in the Earth Sciences Sector of Natural Resources Canada. This paper will discuss these approaches.

The Scalable Vector Graphics (SVG) format is most suitable to create dynamic web-based visualizations in the Internet environment. The functionality of SVG for interface development and display of geo-spatial information is based on user selected attributes, from distributed, web-enabled, data services will be examined.

The selected illustrations intend to explore interface design that uses Scalable Vector Graphics in an application that presents urban land use data.

2 Positioning_our_work

A thoughtfully crafted map offers information as a complex of understandings rather than a monologue of absolute facts. The possibility of the map to be an enriched field of user exploration, inquiry and intent depends entirely on thoughtful, carefully considered design that seeks to maximize the variability of possible understandings. When fully realized, the interface is an arena to hybridize data, interface and information display. It invites the user to work toward a trans-figurative position, to find new contexts and ways of mixing the imagined and the hypothesized. It is dependent upon creating a space for agency rather than one of unilateral transference, upon staging the information for consideration rather than enclosing it, upon the production of variability rather than the illusory pursuit of received truth.

Our work moves outward from an Activity-Centered Design approach where activities are comprised of tasks which are comprised of operations *(http://www.jnd.org/dn.mss/human-centered.html)*. We relied on its focus on "work to be accomplished with the interface" and its hierarchical logic as a way to ground the way we problematized each emerging design question, rather than pinning our work on it as a totalizing theory. Also, working from the idea of Activity-Centered Design displaces attempts to produce a privileged user group. The interface opens up many points of access to maintain a universal and democratic position.

This paper is about how we have engaged with the design questions presented to us, and the work we have produced. The solutions at which we have arrived are not singularly useful to any one application. In fact, our work has been energized by the experience of so many other designers. Little of what is here is ours alone; we have learned from the world wide wide (*zefrank.com* and *yugop.com* both are richly inspirational), and an endless list of material and computational interface environments. We offer this paper as our design statement and hope our experience will be useful to other map designers and interface designers as they search for solutions to their own design questions.

3 Dancing_without_gravity

Design for the Ottawa Land Inventory Map module grew out of a loosely expressed and often shifting list of needs. Scope of work, audience, functions, land use types to be included, all were redefined throughout the life cycle of the design. This necessarily meant we had to adopt an approach that "over shot" any stated goals when formulating our design brief. Although this strategy demanded greater design output, the output could be edited to respond to a variety of anticipated needs. Additionally, the strategy forced deeper examinations of possibilities that constitute a record of valuable design research.

A constant in our mandate was that we should implement our work through Scaleable Vector Graphics (SVG). SVG is an open, vendor neutral, XML (Extensible Markup Language) based vector graphics standard for the depiction of resolution and device independent two-dimensional vector and raster data. At present SVG is a W3C recommendation (w3c.org). SVG supports animation though start time, duration, and attribute values. It implements animation by supporting the modification of graphic objects colour values, coordinate values, and transformation values. It also permits motion of an object along a path.

4 Our_approach

To design an interface for the Ottawa Land Use Inventory Map we found that we needed to extend our consideration beyond the singular program of the map to the relationships presented in the wider interactive field. This lead to an investigation of the information space as a formal typology. By problematizing the information space as a formal typology, elemental structures are free to assume distinct yet co-ordinated roles. These roles are articulated through a precise graphic vocabulary which is directly responsive to user events.

We developed the work as a cohesive, understandable product that mirrors the intentions of the user rather than that of the author. Important to achieving this is the effort to produce neutral space. Without deconstructing an already turbulent discourse, an examination of the idea of spatial neutrality informed an intention that was an underpinning of how the work was art-directed. The aesthetic propensities of the art director will surface; it would be disingenuous to suggest differently. However, the construction of the "neutral field" is used as a strategy to open up the possibilities within the information space by staging the information for consideration rather than enclosing it. This provides the best possibility for user intention to direct how the work will unfold.

We considered ways to create integration of various indicators—in maps, graphs, text and image—while maintaining a graphic signature for each theme. Identifying graphic signatures assures the user still can clearly read the individual themes even while information from two or more sources is juxtaposed in graphing displays. The work was informed by digital form and space, merging of information and display typologies, linked action/reaction interface repetoires and a search for gestural operations.

Within the interactive field, we located projected points of coincidence, inertia, likeness/difference, and imagined how to maximize these potentials. The strategy was one of anticipation of to-be-delivered information and insertion of interface/information programmes. These anticipations were the result of a formal critique of programme intent. Parallels were recognized between the dialogue of use/intent and graphic response/shape, and the already established dialogue of programme insertion and orientation to site. Information display and information manipulation have been collapsed into simple graphic gestures that will accept a user's gaze, focus and consideration.

5 The_work

Basic graphic forms were explored in studies for a text menu window design, as it is the ubiquitous frame for all information in our interface design. The evolution of the window design is explained in these studies.

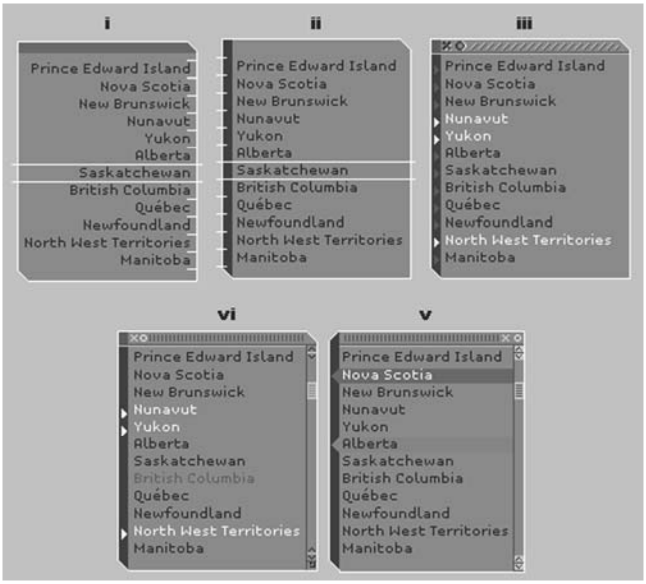


Fig. 1. Text menu window design

i — A coloured bar tops the menu window. This stabilizes the form and becomes user affordance for dragging and repositioning the window. Angled corners adds tension and visual interest to the form.
ii — The coloured bar has been reoriented and now precedes the menu items, providing a directional field for cursor movement.
iii — A textured dragger bar expresses user affordance more clearly. "Collapse" and "close" controls are introduced. Now placed vertically, the colour bar accepts user feedback/cursor tracker responsibilities, making it a focus of user interactivity for the window.
iv — The dragger bar is given a less heavy-handed texture borrowed from Sun JAVA's flush 3D effect *(http://java.sun.com/products/jlf/).* The "collapse" and "close" controls are redrawn to reflect this new subtlety. A

scroll bar is added, with an “expand window” control at the bottom. Cursor tracking migrates to typography, with colour delivering the feedback message. This cleans the colour bar of the visually intense tracking display. The result is a greater focus on selected items, with less visual clutter.
v — The tracking devices of the vertical colour bar have been extruded into ribbon forms extending across the type field. Tracking and selection functions are signaled through colour shifts of the ribbon. A reversed out menu element reinforces the visual cue of the selected state.

5.1 Art_direction

The colour story has been taken out of the palette of 216 web safe colours and is maintained throughout the work. This insures cross-platform consistency; although with current monitor colour depth and resolution, that requirement is vestigial. It is an art director imposed restraint to build a tightly composed working vocabulary.

Typography is executed with Unibody-8 *(www.underware.nl)*, a type family designed specifically for screen use and to maintain its integrity on both Mac and PC operating systems. As it contains roman, italic, bold, black, and small caps fonts, it is possible to set type for every need in the interface.

These carefully applied restraints collect all interface elements in easily recognized relationships that remove former vernaculars while celebrating new graphic languages.

5.2 Map_design

A map of Canada serves as the starting point for a user’s examination of the urban land use indicators (illustration viii). An oblique orthographic projection was selected for the base map. This reduces the northern land mass as it curves over the horizon, and focuses the user's attention on the cities in the southern portion of the country. It is a dynamic effect with a sense of depth and movement. Land mass and water bodies are rendered in an intentionally muted palette of neutral mid-tones and visual clutter is minimized by removing shore lines. In this context, boundaries are useful references but not important information, so they have become undifferentiated white lines to merge with the negative space. The base recedes allowing the point symbols, typography and indicator numbers to be read without competition. Great care was taken in the execution of typography for this map. A subtractive strategy was used in this design to ensure maximum clarity of type and indicator numbers.

A window listing indicators and available years of Canada Land Inventory Land Use data (illustration vi) is the entrance and circulation space for user investigation. From here, indicators may be selected to be shown on roll over (the squares in front of the list are check boxes) on a city map of Canada, as well as being graphed as histograms (illustration ix). Illustration viii shows a rollover of Winnipeg as it would appear if the use had turned on all indicators; and a rollover of Ottawa if only "Compactness" was turned on.

The histograms and numeric lists of indicator values are dynamically generated SVG data views. Indicator values are organized in a cities list window which can be collapsed and closed, letting the user control a potentially visually intense display (illustration ix). By respecting human cognitive processes, complexity is elegantly managed through this strategy of window control while continuing to reveal the structure of how the data are organized.

6 Ways_of_looking

When the indicator numbers were listed with the city names, certain possibilities became evident. We offered sorting methods based on ascending or descending values of indicators, east/west, west/east location of cities, as well as alphabetically by city name (illustration ix). As the data are manipulated, the histograms are also re-ordered. Each sorting method has the potential to reveal a fresh look at the nature of the data, letting the user understand its structure in new and useful ways. It does not require a 3D fly-over, or an animated sequence. Nor does it require a map. These discoveries happen in a simple list of city names paired with statistical values. This seems to be closer to the essential idea of visualization as being about ways of looking, rather than being necessarily bound to the latest, trendy widget.

Urban land use is a subject matter that is necessarily technically dense. To lessen this esoteric situation, a glossary of definitions for indicators and land cover types is proposed. The definition is presented in a window which appears when the user clicks on an indicator name from the main selection window or a land cover entry from the legend. See illustrations vi, vii (and x).

The national module is an avenue into a city scale study of urban land use. At any time in the user's exploration, the user may leave this national scale module and move to an urban scale view of the data by clicking on a city from either the map or the list of city names. (See illustrations viii and ix.)

Through static and dynamically generated hyperlinks, SVG permits users access to more detailed views of indicator values. This is the means to allow users to move from the national view to an urban view as well as to plot graphs and histograms of layer values already selected in the legend. Scaleable Vector Graphics may also be used to dynamically generate attribute information analysis through animating these histograms and graphs.

Spatio-temporal analysis may be performed using SVG to dynamically access web map server and web feature server data to select a series of temporal moments so as to detect change over time.

Fig. 2. Main control window and map view

vi — This view of the main control window shows several states: the selected year is 1971 and the “Compactness” indicator is on.
vii — This explanation of “Compactness” is revealed by clicking on the indicator name.
viii — The national view map showing a roll over of Ottawa with the “Compactness” indicator selected, and a roll over of Winnipeg as it would appear with all the indicators on.

ix

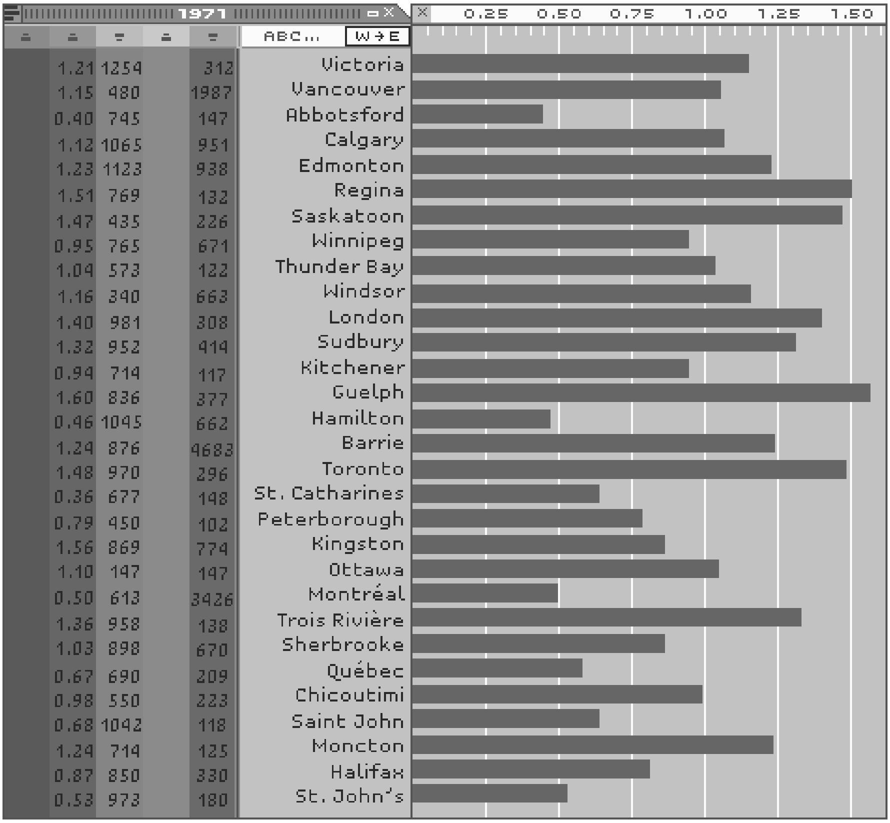


Fig. 3. Histogram window

ix — The cities list with the histogram window open. We see three indicators are turned on, compactness is plotted and the list is sorted by city from west to east.

7 Interface/information collapse

Interface/information collapse is evident in the overview map/navigation device/scale control (illustration xi). This combines two interface devices in transparent, gestured operations available in one location. User affordance is provided through a cursor change when rolling over the device. When the pointer is inside perimeter of the rectangle, it is replaced by the often-seen dragger hand. On mouse down, the user can drag to

reposition the rectangle, and so reorient the map inside the neat line. Our improvement to the scaling operation is a combination of enlargement and reduction in a simple appliance: a slider bar. When the control/indicator is moved to right, the scale is increased. When it is moved to the left, the scale is reduced. Needed understandings of relative scale are read from the ratio of the map outline:entire overview map. (Please see the navigator window in Adobe Photoshop version 5.0 and the
panning device implemented at *yoox.com* and *earth.google.com*.)

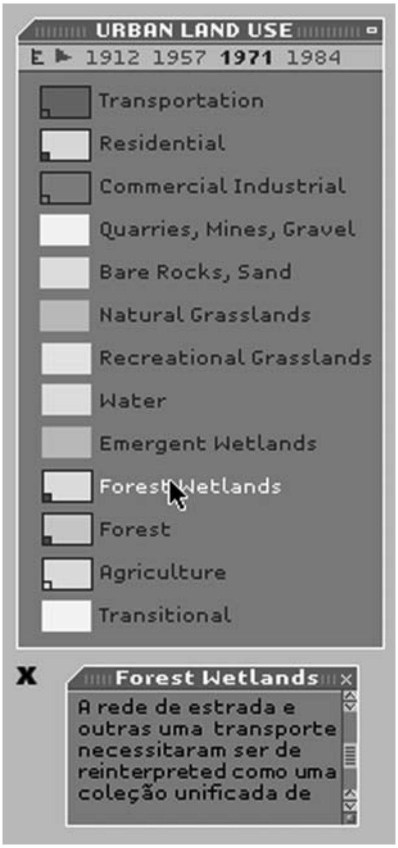


Fig. 4. Legend

x — The legend is the year selector/layer control for the map. It also lets the user make active other data display types, as well as playing an animation which cycles through the years. Clicking on the legend's colour blocks turns layers on or off. Infotips roll overs are managed through the square inside the "on" layer's colour block. In this screen capture, data for 1971 is being explored. The transportation layer is on, but roll overs for the layer is off. Roll overs is turned on for the residential layer. Forest wetlands is highlighted (user feedback) before it is clicked to open a window giving its definition.

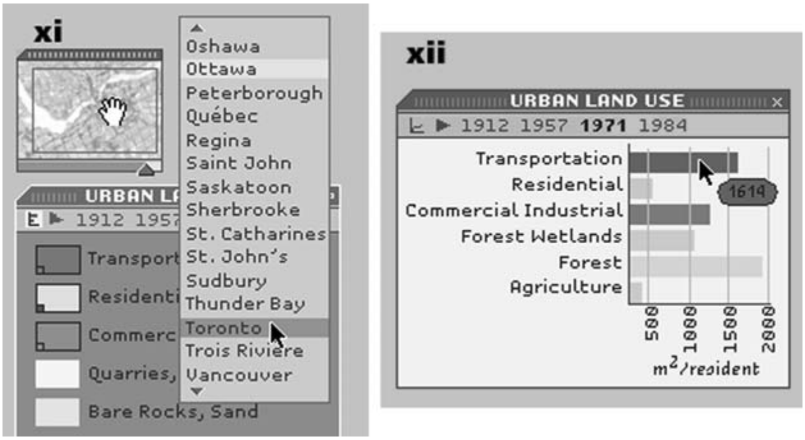


Fig. 5. Scaling/Navigation Device and plot of histograms

xi — The scaling/navigation device showing the dragger hand. The popup scrollable menu of city names. Ottawa is the current city being examined and the new selection will be Toronto.
xii — A plot of histograms of the currently selected layers. An exact value for "transportation" is given as a roll over. Year selection, animation through data years and a graphing window are all available from the top of this window.

Fig. 6. Map interface elements

xiii — A composition of map interface elements. The scale/navigation device and city selector device are permanent elements. In this illustration, the window of graphed land cover values also takes advantage of roll overs to give exact values of each plotted point. The indicator selection window affords examination of other land use themes.

A small map of Canada is paired with a popup menu of city names. This provides a means for the user to select a new city or to return to the national view of cities and indicators.

Function/information coincidence is best illustrated in the legend window. Roll over control, legend, map item on/off control are collected in one collapsible window. The window-as-interface is further developed by including "year selection" and "show histogram" controls at top. The legend's essential role in traditional map design (that of typological clarifier) is taken up and advanced in our design.

User intention is carried across interface/information devices. Those layers "turned on" in the legend window will be shown in the histogram and graph windows (illustration xiii). Standard SVG roll over functions are explored in these graphing windows. An overview and general trends are available to the user from the graphs, with exact values given when the user rolls over a histogram or a graph node (illustrations xii and xiii).

8 Conclusion

What we set out to accomplish in our work to design this interface is to give the user a sense of control and understanding. We want the user to see the purposes of the interface and its reactions to user events. In turn, we want that to smooth the way for the user to understand the information that is available.

"Weightless" is an adjective we kept in mind during this dance. We have placed our focus on the idea of a weightless experience that does not feel like work, but feels like a natural extension of a user's intensions. This allowed us to substitute simplicity for complexity and to propose an interface design that offers ease of use, and thoughtful ways to engage the information.

This chapter is our design statement. We hope our experience will be useful to others as they search for solutions to their own design questions.

Landmarks and the perception of a space in web-delivered 3D-worlds

William Cartwright

School of Mathematical and Geospatial Sciences
RMIT University, GPO Box 2476V, Melbourne, Victoria, 3001
william.cartwright@rmit.edu.au

Abstract

This chapter reports on a research programme that is building and testing a 3D tool for Community Collaborative decision-making. It concentrates on the stage of the research that focuses-on the question: How does a priori knowledge of an area change navigation and exploration abilities when ‘moving through’ and exploring the Virtual World? That is, how is the perception of a place, built through the use of a virtual environment, modified by prior knowledge of an area being studied / explored? As well, does this effect a user’s ability to explore a virtual geography more effectively? We wish to ascertain how the users’ prior knowledge about the real world can be applied to building more effective virtual environments.

From naive user to expert user we wish to find-out what elements need to be ‘built’ into the model to support wayfinding, exploration and subsequent data visualization and decision-making. A priori knowledge and previous experience and methods of map and other geographical information artefact usage will also be considered, as it will provide important clues about ‘preferred’ methods of tool use.

The chapter will provide a description of the project and its potential use in an urban planning context, briefly provide the results of previous research that form part of the project and provide results regarding how

users perceive an urban space through the use of a combination of landmarks and simple ‘block’ building shapes.

1 Introduction

There exists a difference in how users locate themselves in the context of today’s digital electronically connected world. And, it is argued; that perhaps how users locate themselves is different, depending upon the artifact used. When using Location-Based Services (LBS) or ‘At-Location Mapping’ (ALM) (where maps are served to users ‘at location’, providing them with relevant and timely artifacts with which to navigate or explore). While there has already been considerable experimental work conducted using some of these new technologies to convey geographic information to users, we are still unable to say with any firm measure of confidence whether images delivered on small mobile devices need to still contain all of the information contained in paper maps. Accordingly, rigorous evaluation programmes must be designed and implemented to understand the strengths and limitations associated with these new multimedia mapping devices and apply them more effectively and efficiently in ways that improve maps displayed and map use. This issue is of critical importance given the massive uptake of mobile devices and information technology in general.

2 Overview

A project is underway to supplement the community decision-making process with a World Wide Web (Web)-delivered interactive 3D tool. This involves building and testing a 3D tool for Community Collaborative decision-making. A research team at RMIT is building virtual 3D audiovisual models of urban spaces to test their potential to enhance community discussions of future neighbourhood developments. Of great interest to the team is how successful these simulations are, especially when delivered through the sometimes restricted ‘pipe’ of the Internet. Also of interest is how these tools are accepted by the user group and how the tools might best be designed and delivered to suit community use.

A VRML world was created for the study area. The VRML world is shown in figure 1.

Fig. 1. VRML world

The world was developed as a test bed to determine 1. What the model should contain – a trade-off between development costs and usability, 2. How much information needs to be included for professional and public users and 3. How landmarks might be incorporated to facilitate a ‘balance’ between minimal information provision and usability.

An initial evaluation of the tools was made with a focus group to provide general feedback. A second evaluation was conducted so as to better understand how complex a computer graphics 3D environment needed to be for community discussion of urban planning developments. This chapter reports on the last Stage of an evaluation project which was undertaken to ascertain how landmarks and the users’ familiarity of a study area can be used to build simpler and thus more economical virtual worlds.

A problem for developers of decision support tools is to provide adequate assistance to all stakeholders in the decision-making process. Tools must be developed for expert and non-expert alike. In inner urban areas, where a plethora of building types abound, and historical knowledge is hidebound to understanding how an area ‘works’, it is important to provide visualizations that support the visualization process. However, many of those involved in the decision-making process may only have a very naïve understanding of an area. For these users perhaps a more detailed model is needed, and for those with a greater knowledge of an area the model could provide less detail. The focus of this evaluation was to determine the level of detail that would be ‘acceptable’ for a 3D urban visualization tool.

If users cannot recognise where they are they will endure stress, and search for geographical information that puts things in (spatial) perspective (eg finding their bearings, orientation to north etc.). According to Golledge (2000, p. 1) “We often assume there is no need to learn this type of geography because we already “know” it! And, we have not bothered to make this underlying geography explicit. Golledge thinks that naïve geography gives an implicit knowledge via environmental perception and that landmark or feature recognition and an awareness of the built environment is part of geographical understanding. He says that “People who claim they ‘can’t do’ geography can provide accurate assessments of their local area. …“Users ‘already know it” (Golledge, 2000, p. 1). For example, aspects of the geography of daily life that we “implicitly “know” but have not bothered to make the underlying geography explicit (*ibid*, p. 7). Naïve geography gives an implicit knowledge via environmental perception, through the use of landmark or feature recognition (*op cit*.).

Navigating in the real world can be difficult for users who need to travel through unfamiliar terrain. This difficulty is enlarged when they have little local knowledge of an area as well. When users have to navigate in virtual worlds the problems can be different, but still certain ‘knowledge’ of landscapes could assist with wayfinding in this synthetic reality. So, what happens when inexperienced users are ‘plunged’ into a virtual world, and asked to explore and understand a geography they may not properly understand? And, how do they best navigate through this world? When users wayfind and explore in Virtual Worlds severe problems do exist (Darken and Sibert, 1996). Without cues users can become disoriented and may not be able to properly use a virtual environment package.

According to Darken and Sibert, wayfinding strategies and behaviours are strongly influenced by environmental cues. In the type of environment that is being simulated – an inner urban environment, consisting of a diverse range of building types and different densities – the environmental cues are mainly buildings, with few other cues or landmarks to assist location determination and wayfinding (in the virtual world). Knowing about an environment prior to undertaking a study using a virtual environment can make it easier to interpret n urban simulation. Conversely, little knowledge of an area can make it very difficult to visualize a real space by using a synthesis of it. Therefore, does a done model need to be provided for a user who is knowledgeable about an area and another to someone unfamiliar with that area?

What this stage of the research focusses-on is the question: How does a priori knowledge of an area change navigation and exploration abilities when ‘moving through’ and exploring the Virtual World? That is, how is the perception of a place, built through the use of a virtual environment, modified by prior knowledge of an area being studied / explored? As well, does this effect a user's ability to explore a virtual geography more effectively? We wish to ascertain how the users' prior knowledge about the real world can be applied to building more effective virtual environments.

From naive user to expert user we wish to find-out what elements need to be 'built' into the model to support wayfinding, exploration and subsequent data visualization and decision-making. A priori knowledge and previous experience and methods of map and other geographical information artefact usage will also be considered, as it will provide important clues about 'preferred' methods of tool use.

We are interested in ascertaining the users' Spatial Knowledge (Darken and Sibert, 1996), which is based on survey knowledge – the ability to conceptualise a space as a whole, enhancing wayfinding skills (Thorndyke and Goldin, 1983, cited in Darken and Sibert, 1996). This is different from procedural knowledge (defined as the sequence of actions required to follow a particular route). It can be acquired from map use and it is also based on landmark knowledge (static information about the visual details of a specific location (Darken and Sibert, 1996).

The evaluation was based on the three primary wayfinding tasks specified by Darken and Sibert (1996). These are:

- Naïve search – where the navigator has no a priori knowledge of the whereabouts of the target, requiring an exhaustive search;
- Primed search – where the navigator knows the location of the target, performing a non-exhaustive search; and
- Exploration – where there is no target.

The naïve search, as Darken and Sibert note, is rare in the real world, but common with first-time users of a virtual space. Naive search will therefore rely on certain wayfinding aids to support their movement through the virtual world. Darken and Sibert (1996, p. 4) proposed that the basic principles of organising an environment to support wayfinding were to:

1. Divide the large-scale world into small, distinct small parts, preserving a sense of 'place';
2. Organise the small parts under a simple organisational principle; and
3. Provide frequent directional cues.

They also recommended that map design principles should be applied to the model, providing an orientation-independent representation of the environment. This entails:

- Showing all organisational elements, roads, landmarks, districts, etc.) and the organisational principle;
- Indicating the observer's position; and
- Orient the model with regard to the 'forward-up' equivalence of map reading (ie having the 'forward' direction of travel aligned with the map).

The last requirement demands interesting 'cartographic callisthenics', as maps are rotated this way and that, so that the map is oriented with the forward direction of travel, with the final destination being always placed furthermost from the user's body. But, when moving / navigating through virtual worlds this is done automatically as the user is unable to rotate the screen. It is argued that this would impose a greater cognitive load.

This stage of the evaluation process will have as its goal the determination of what users need to know or understand about SPACE (general information about the 'area' being studied) and PLACE (dictated / determined by location and purpose-specific elements that are unique to the particular user and their usage requirements). What needed to be resolved were

- users' concepts of space;
- their concept of place;
- how they navigate through space; and
- how they navigate through their personal place.

3 Evaluation

29 candidates participated in the evaluation. Their age range was 18-25 and all had competent to efficient map use skills. The test candidates were split into two groups: Group 1 - identified as having a priori knowledge of the area. 12 candidates were identified as belonging to this group. Group 2 had relatively no knowledge of the area.. 17 candidates belonged to this group. Then each of these groups was further split into two sub-groups: Group 1a and 1b, Group 2a and 2b.

The Candidates first completed a profile proforma to glean information about their proficiency in map and map-related tool use and also their perceived knowledge of the area.

The session operated thus:

1. Groups 1a and 2a were taken on a 'guided tour' of the area.
2. During this time Groups 1b and 2b undertook the evaluation / feed-back of the VRML model.

This took approximately 1 hour.
Then this process was reversed.

3. Groups 1b and 2b were taken on a ‘guided tour’ of the area.
4. During this time Groups 1a and 2a undertook the evaluation / feedback of the VRML model.

GROUP			
1. Local Knowledge		2. No Local Knowledge	
1a Tour 1	1b Tour 2	2a Tour	2b Tour 2

Fig. 2. Groups

Candidates were asked to conduct two searches of the 3D model of the study area, one as a general ‘exploration’ of the area and the other a task-related search. In the task-related search they were required to find specific buildings that are typical in the study area. These searches will be 1. Naïve (the candidates had no knowledge about where they were located) and 2. Primed (they knew the area after a walking tour prior to the evaluation). Here they were asked to view two ‘virtual tours’ of the study area, one as a general ‘exploration’ of the area and the other a task-related search. In the task-related search users were asked to identify the different building types that are typical in the study area. Here they were also asked to note the buildings that you think are the key, or landmark buildings.

The model of the study area used is shown in figure 3.

Fig. 3. VRML model

Finally, candidates were asked to make general comments at the end of all tours and product evaluations.

4 Results

4.1 Level of detail

Candidates firstly considered whether the amount of detail provided was sufficient for them to understand the general geography of the area. Did it provide sufficient information for them to be able to make informed comments about potential developments in the area?
1. The amount of detail is sufficient: All groups found that there was sufficient detail to understand the area.
2. There are adequate landmarks to assist in orienting oneself. Landmarks in the area, as previously noted are mainly prominent buildings. One of these buildings is depicted in figure 4. All found that they had adequate landmarks to assist, except Group 2a. This group had no prior knowledge of the area, and they had been on a tour of the area prior to undertaking the evaluation. Here, the users identified that even though they had been on a tour, they thought that extra information was required.

Fig. 4. Landmark building

3. Having all buildings in full detail is necessary. All groups except 1a thought that there was sufficient detail. This is interesting, as group 1a were knowledgeable about the area and they had completed a tour prior to

the evaluation. Therefore their comments will need to be further investigated.

4. I could understand the area with less detail in this 3D model, which would provide me with an adequate mental representation of the area.

Group 2b thought that more detail was required. This group had no prior knowledge of the area and they had not been on a tour prior to the evaluation. The other part of the 'no knowledge' group (2a) thought that the level of detail was sufficient, but they did not indicate a full support of the amount of detail. Therefore, when users have little knowledge of the area they will respond better to the model if a tour is conducted prior to actually using the tool.

5. Less information / detail would still allow me to build a mental image of the area. Again, those who had no knowledge of the area could not accept a model with less detail. And, the members of this group who did not undertake the tour indicated that they could not work with less detail. This again supported the concepts that having knowledge of the area allows for a simpler model to be provided, and that a pre-use tour assists in better exploiting the model.

6. Having all elements in full detail makes the image too complex. (ie it has a negative effect, rather than improving the model). All candidates generally disagreed with this statement. The level of detail did not make the tool more complex.

7. I need the addition of street signs for me to orientate myself. Street signs were added to the model after feedback from previous evaluations. A typical street sign is shown in figure 5. The inclusion of the street signs were supported, but less so by groups 1b and 2a. (Later comments about the inclusion of the street signs included: "Street signs also provide human scale in terms of height".

Fig. 5. Street sign

8. ‘End-of-the world’ images make the 3D image look more real. These images were added to the model so that it did not appear to ‘end’ at the edge of the VRML world. All candidates supported the inclusion of end-of-the-world images.

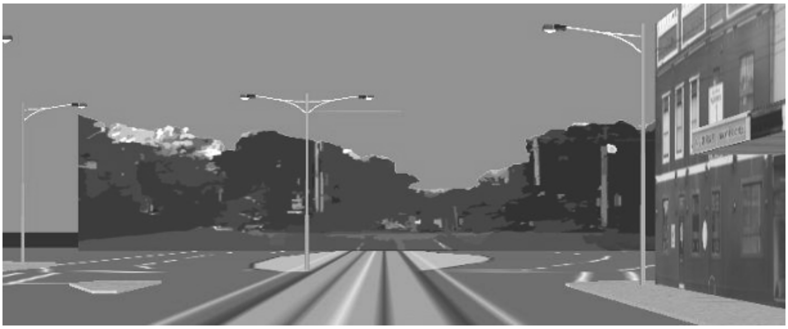


Fig. 6. End-of-world image

9. Adding light poles and wires makes the 3D image look more like an inner Melbourne shopping strip. This inner urban area has the usual trappings of overhead wires, poles and banners. The test prototype provides the option of having these items ‘on’ or ‘off’. The images in figure 7 illustrate this. All candidates thought that the addition of these items was necessary.

Fig. 7. World with and without wires, poles, etc.

10. Changed environmental conditions make the 3D world more appropriate for better visualizing local conditions. The prototype model allows for the environmental conditions to be changed – sunny to overcast, day to night. These items can be chosen by selecting the appropriate radio buttons in the interface. However, the test candidates did not think that the addition of this function enhanced the tool use. These different environmental conditions are shown in figure 8.

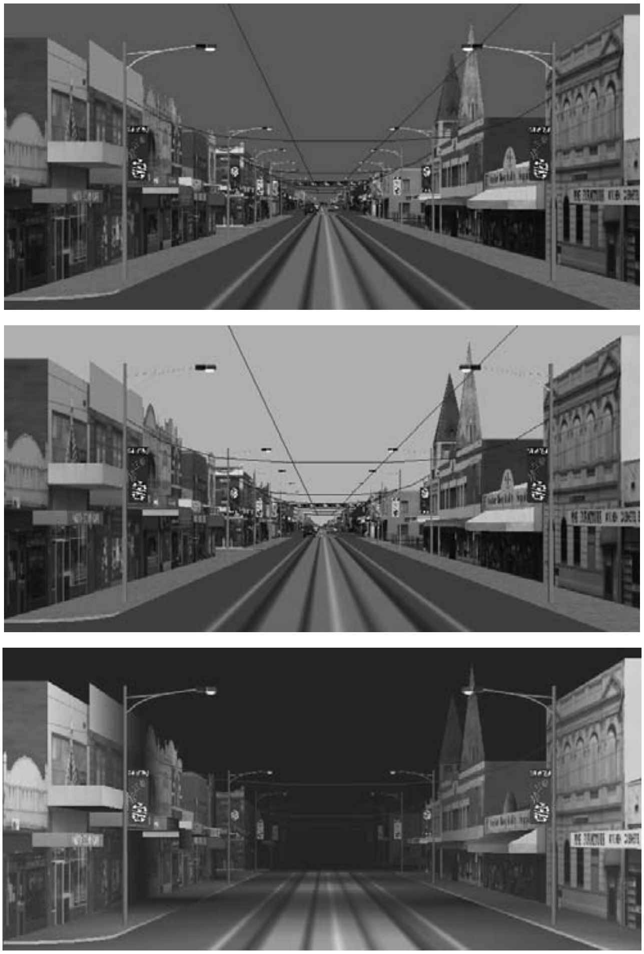


Fig. 8. From top: during the day and sunny, overcast and at night

11. The area consists mainly of small shops. All candidates thought that this was the general concept of the area, both before and after the tour.

Group 2a thought that the area consisted of more than ‘just shops’. This perception of the area was considered in the next question.

12. The area consists of shops and some significant buildings. All candidates agreed with this statement. Therefore, it is thought that all of these elements must be provided in the model.

A summary of the results from the evaluations is provided in the following table (table 1).

Table 1. Summary of questionnaire results.

		Group 1		**Group 2**	
No.	Questions	a	b	a	b
1	The amount of detail is sufficient.	☺	☺	☺	☺
2	There are adequate landmarks to assist me in orienting myself.	☺	☺	😐	☺
3	Having all buildings in full detail is necessary.	☺	😐	😐	😐
4	I could understand the area with less detail in this 3D model, providing me with an adequate mental representation of the area.	☺	😐	😐	☹
5	Less information/detail would still allow me to build a mental image of the area	☺	☺	😐	☹
6	Having all elements in full detail makes the image too complex. (i.e. it has a negative effect, rather than improving the model)	☹	☹	😐	☹
7	I need the addition of street signs for me to orientate myself.	☺	😐	😐	☺
8	'End-of-the-world' images makes the 3D image look more real.	☺	☺	☺	☺
9	Adding light poles and wires makes the 3D image look more like a shopping strip.	☺	☺	😐	☺
10	Changed environmental conditions makes the 3D world more appropriate for better visualizing local conditions.	😐	☺	😐	😐
11	The area consists mainly of small shops.	☺	☺	😐	☺
12	The area consists of shops and some significant buildings	☺	☺	☺	☺

4.2 Landmarks

Candidates were asked to identify what they thought were the landmark buildings in the area. That is, if only some buildings could be shown in full detail and others in outline mode only, which buildings must remain in full detail to allow you to properly navigate through the area. As well, these 'landmark' buildings, plus all other buildings as outlines would enable candidates to build a 'mental image of the area. They were asked to consider that the model must provide them with sufficient information for them to be able to make informed comments about potential developments in the area.

Landmark recall was best for those groups who undertook the tour, and worst for the group with little local knowledge and no tour. Candidates indicated that all landmark buildings must remain in the model.

4.3 Inclusions

The final section was a general discussion with the entire group. Here we wanted to find out what the best combination of the world might be. We wished to ascertain which additional world elements are appropriate for inclusion in a model that consists of only landmarks, outline only secondary buildings and additional elements.

During this part of the evaluation candidates were asked as a group to decide which elements should be included – using the ON/OFF buttons that the demonstrator clicked. These buttons are right of the Web page shown in figure 3. An enlarged section of this figure is shown in figure below:

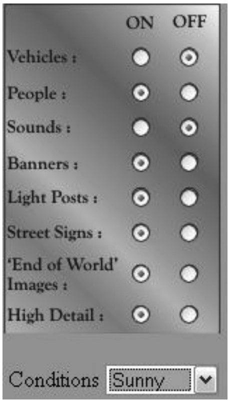


Fig. 9. Control panel

In addition, candidates were asked to provide information about the combination of items that they thought should be included. This was a combination of the on/off elements shown in figure 9.

Those with a local knowledge of the area wanted traffic to be added and that street signs were necessary. Those with no local knowledge liked the night/day function and they thought that the street signs were useful. They indicated that perhaps the model would be improved if we were to add signs to specific features, like the railway station. They also thought that the model needs cars in the street and landmark buildings. They generally thought that the more detail the better.

General comments were also solicited from the group. Their comments are summarized below

Group 1 – Local Knowledge

- If detail could be added, it should, as it improves the understanding of the area.
- As the area has heavy traffic, this should be added.
- Street signs are an excellent inclusion.
- Less detail means that the ‘feel’ of the environment is lost.

Group 2 – no local knowledge.

- Night/day function useful.
- Street signs useful.
- Perhaps add signs to specific features, like the railway station.
- Needs cars in the street.
- Needs landmark buildings.
- The more detail the better.

And general comments from both groups:

- Include end of the world images
- Landmarks need to be left at street corners.
- If only landmarks are shown, then the character of the area is lost.

5 Relevance of research results to LBS and Small-screen devices

Whilst the research reported upon in this chapter focused on the importance of landmarks in 3D VRML worlds the results obtained can be used to inform researchers developing other visualizations for navigation and interpretation. The need to show all content in 3D was not evident from the results obtained. It was found that landmarks were only needed for significant buildings and surrounding buildings could be shown as simple outlines.

When developing LBS and other services via the mobile Internet this is an important design consideration, as if all buildings do not have to be rendered then transmission rates would be improved dramatically. However, this research needs to be applied to small, mobile devices to better understand if the results from the research can automatically be migrated to LBS and TeleCartography applications.

6 Conclusion

This research project has evaluation 3D models for community collaborate decision-making support in three stages:

1. An initial qualitative evaluation of an Alpha product with an expert group of users;
2. Testing how the 'geographical dirtiness' of the Virtual Environment changes the perception of a space; and
3. Discovering the appropriate wayfinding aids needed in the model to support searching and exploration.

This chapter has reported on the last of these stages. It focussed-on two questions: 1. What is the minimum number of landmark buildings that should be included in the model so that it provides adequate information about the area? And, how does a priori knowledge of an area change navigation and exploration abilities when 'moving through' and exploring the Virtual World? That is, how is the perception of a place, built through the use of a virtual environment, modified by prior knowledge of an area being studied / explored? The study was done with two evaluation groups – one with knowledge of the area and another without this knowledge. As well, tours of the area were conducted to ascertain the trade-off between time taken to tour the area and the time involved in further enhancing the model.

The results reported in this chapter will be used to further enhance 3D models and to make them more useful for navigation and comprehension.

References

Cartwright, W. E., Pettit, C., Nelson, A. and Berry, M., 2004, "Building community collaborative decision-making tools based on the concepts of naive geography", GIScience 2004, Washington, D. C., USA: Association of American Geographers.

Darken, R. P. and Sibert, J. L., 1996, "Wayfinding Strategies and Behaviours in Large Virtual Worlds, *CHI96 Electronic Proceedings*, http://www.acm.org/sigchi/chi96/proceedings/papers/Darken/Rpd_txt.htm, Web page accessed November 1, 2000.

Golledge, R., 2000, "Geography and Everyday Life", *Directions Magazine*, http://www.directionsmag.com/columns.php?column_id=37, 14pp, Web page accessed 14 May 2004.

Thorndyke, P. W. and Goldin, S. E., 1983, "Spatial Learning and Reasoning Skills", in: *Spatial Orientation: Theory, Research and Application*, H. L. Pick and L. P. Acredolo (eds.), New York: Plenum Press, pp. 195 – 217.

Comparing the effects of different 3D representations on human wayfinding

Michael Wood[1], David G. Pearson[2], Colin Calder[3]

[1] Department of Geography and Environment, University of Aberdeen
[2] School of Psychology, University of Aberdeen
[3] Learning Technology Unit, University of Aberdeen
geo125@abdn.ac.uk

Abstract

This paper reports on a preliminary stage of some new research comparing the effectiveness of conventional topographic maps and computer-based geovisualisation systems as aids to navigation in wilderness mountain areas in Scotland. The whole programme considers both the cartographic products and the perceptual and mental processes involved in their use but this paper concentrates on some fundamental issues and the results of the first experiment which trialled main procedures and concepts. In both this and the main experiments participants were asked to judge gradients and journey times for various routes presented on maps and interactive animated 3D terrain displays. The first experiment – reported here - compared only terrain models with different surface rendering, and produced unexpected results. Subsequent studies[1] included the essential comparative component of the 2D map. When complete this research programme will provide an important evidence base for future mapping system producers seeking to develop generally accessible terrain visualization tools.

---

[1] in progress

1 Introduction

The region of the hippocampus of rats (Hafting, *et al,* 2005) has recently been identified as showing evidence of tangible cognitive maps, exhibiting topographically organized spatial tessellated structures anchored to external landmarks. Although the full relevance of these findings to humans has yet to emerge there is increasing recognition that the internal (cognitive) representation of spatial environmental knowledge is not only fundamental to humans but essential for spatial decision-making, especially in the context of wayfinding. Most perceptual information employed in wayfinding activities is obtained from direct environmental experience. However, since the emergence of what has been called the writing-based ‘theoretic culture’ (since c. 45 000 BP) and the associated invention and development of visiographics and external memory (Donald, 1991), external facilities such as maps and other related images have become increasingly common as aids to thinking and are believed to influence both the nature and content of cognitive maps and the wayfinding strategies which result from their use.

Traditional paper maps have been valuable sources of geospatial information for travellers, both for planning and on-going navigation, but complex environments lead to the increasing graphic complexity of such products which has challenged both map makers and users. Good cartographic design and appropriate training in map use have helped to accommodate such complexity, but the creative application of new computer-based technologies can now add significantly to environmental knowledge acquisition and help enhance decision making processes. Today, despite the increasing incorporation of computer and communication technologies into daily life, customised paper maps still exist alongside navigation aids such as Memory-Map (URL 1) and Anquet Maps (URL 2), which provide users with some interesting alternatives. Printed topographic maps have long been the tools of choice for route finding in mountains, but continue to demand considerable perceptual and cognitive effort for their interpretation and use. Understanding elevation contour patterns requires special knowledge of symbol systems, the landforms depicted and the navigation task. It has been reported that orienteers, for instance, develop, from preliminary study of the map, a view of how parts of the terrain will appear. This "pre-field-experience," impression derived from the map is later refined and attuned to the real terrain (Ottosson, 1996). Some experienced orienteers can develop and maintain, while running, a mental ‘map model’ (a form of cognitive map) of the competition area. Some retain this over time and refer to it intermittently without the need to look back at the original physical map. This can give them considerable competitive advantage over those who must keep pausing to check with the map to strengthen their mental image[2]. As the expertise required for such mental modelling may involve the user in imagining the interpreted landscape as a 3D mental image, the current study is investigating the potential of new geovisualization techniques (compared to traditional maps) to augment these abilities. If screen-based computer generated models, resembling those derived mentally from contour maps by some experts, can now be created and viewed interactively on computer screens, less experienced users may also be able to acquire some of the advantages of those with greater skill and training in traditional map reading and more experience of mountain route-finding. This follows the logic of using models in many complex situations to help develop awareness of the real objects and environments.

The wider research programme, of which this paper forms part, is also seeking answers to psychological questions about the nature of the interaction between conventional and computer-based products and mental representations – largely ignored by current theories of human working memory[3]. Put simply, if mental representations from experience of reality are many times better than those gained from reading static 2D maps, does learning from dynamic 3D models give results somewhere in between? The study is also, simultaneously, addressing some of the 'research challenges' - specifically cognitive/usability issues - of geovisualization tools (MacEachren and Kraak, 2001). Separate experiments examine the relative effectiveness of interactive, animated 3D terrain models, and traditional contour maps as external aids for the tasks of estimating the slope gradients of mountain paths and the time taken to complete predetermined mountain walking routes. Recognition memory for 2D and 3D representations is also being explored.

The experiment reported in this paper compared only representations of terrain models. The unexpected outcome was that, despite considerable differences in surface rendering, the two models employed proved to be equally effective in the context of the experimental tasks.

---

[2] The sport of orienteering permits only the basic tools of map and compass to help navigate between checkpoints along an unfamiliar course. Modern aids such as GPS and other Location-based services are illegal.

[3] The part of the brain that provides temporary storage and allows manipulation of the information being used in complex cognitive tasks such as learning, and reasoning.

2 Some preliminary studies

Although a genre of pictorial maps and panoramas, often of high artistic quality, has emerged during recent centuries (Wood, 2000), few have satisfied the more rigorous characteristics of both reliable consistency and visual ‘friendliness’ required by unskilled map readers to support spatial reasoning tasks such as detailed wayfinding. By the 1980s digital terrain modelling software had been developed for the production of fixed views, but not until the 1990s was it sufficiently user friendly or could it offer visually acceptable output. Pilot studies with two such earlier systems – LaserScan’s Horizon and ESRI’s Arc/Info - laid some of the groundwork for the present investigation. In one case 3D terrain images were created to support map reading experiments with geography undergraduates (Wood & McCrorie, 1993). Results suggested that such views *could* help with contour reading, but a more significant finding was of the considerable difficulty encountered by many participants when trying to interpret contour patterns from the maps themselves, reflecting very limited map based knowledge or experience. If this condition proved widespread, the need for new and more powerful visual tools would increase. In another study (Wood & Goodwin, 1995) static 3D views were again created, this time using ESRI’s Arc/Info, with draped satellite images and overlaid with selected vectors (contours, footpaths and streams). Test participants were grouped by level of knowledge and experience of map reading and mountaineering into ‘real experts’ (serious amateurs, rescue team members), ‘regular leisure walkers’, ‘occasional walkers’, and inexperienced ‘opportunists’ (such as motor tourists who reach a mountain car park in good weather and decide to take a walk!). In the early 1990s a smaller pilot study with real experts (serious mountaineers) led to the following observations about the potential of the digitally generated but more limited symbolically enhanced static 3D views which could be created at that time:

- Useful for route planning but less suited to ongoing navigation *(e.g. through lack of detail).*
- With more realistic surface representation such views might provide sufficient information without the support of a map *(e.g. with a full vector drape of topographic symbology).*
- Some recognition of the possibility that the increased availability of higher quality (and perhaps more realistic) representations could benefit mountain safety *(although such images were not available to these participants at that time).*

New generation terrain modelling software offers much greater potential for the support of external cognition, including better surface rendering, interactivity and the possibility of animation. Output options ranging from orthographic 2D maps to 3D terrain models viewed from any angle and incorporating rotation, zooming and ‘flythro’ capability, provided the flexibility required for the current research programme.

3 The current project: Background and experimental strategy

The increasing availability of computer-based visualisation and multimedia technologies to help supplement or even replace conventional printed maps, has been partly driven by the general assumption that dynamic 3D representations can provide more effective support than 2D maps alone. Few studies, however, have examined whether this assumption is correct (Scaife and Rogers, 1996). Most previous findings on perception and cognition in wayfinding are based on static paper maps alone, and much less is known about issues associated with 3D and especially dynamic displays (Slocum et al, 2001). It is well documented that mental representation of environments stimulated by 2D maps is quantifiably different from representations produced by direct experience (e.g. Thorndyke and Hayes-Roth, 1982; Moeser, 1988). It is therefore important to investigate whether learning, using 3D terrain models (which more closely approximate to reality than maps do) produces different or superior memory than learning using 2D maps of equivalent geospatial data. Another neglected area of research is into the effects of expertise on the use of 2D and 3D representations. Recent theories of expert cognition in orienteering (e.g. Eccles et al., 2002) are based solely on the use of 2D maps. The majority of previous studies have largely ignored direct comparison of 2D and 3D representations, and focused on urban/manmade settings rather than wilderness environments, or on the development and application of the technology for implementing such representations for prospective users (Moore et al., 1999; Morrison and Purves, 2001). The social, economic and environmental importance of mountain regions in many countries have been steadily increasing over recent years, with an associated increase in the use of such areas for recreational activities. Also the users who regularly visit areas such as the Scottish Cairngorm Mountains range in experience from serious mountaineers to casual weekend walkers, many of whom could benefit from additional navigational support. The current project aims to establish a source of data (based on scientifically rigorous investigation) on whether or not interactive, animated 3D terrain models do provide better cognitive support than traditional static 2D maps. This could contribute to better evaluation of the potential impact of such geovisualisation tools for recreational use in mountain environments. The introductory phase of the programme comprised two experiments. The first, the pilot study, was primarily used to trial concepts, materials and procedures. Subsequent experiments focused on the interpretation of both 2D and 3D geospatial data. The remainder of this paper concentrates on the first experiment.

3.1 The stimuli used in the experiments

To provide some background, this section reviews the map-related material devised for the whole experimental programme. Three stimuli were prepared, one topographic map and two versions of a terrain model of the same area. For convenience to the experimental procedures the maps were scanned and presented onscreen along with the digital terrain models. Using ESRI ArcGIS/ArcMap/ArcScene software, models were created of twelve separate 10 km^2 tiles selected from different high level mountain regions in Scotland. Ordnance Survey (OS) 1:25 000 scale digital data were used. The two versions of the models to be used in the experiments had shading from oblique illumination and one, in addition, had a draped image – the raster version of the 1:50 000 OS map. The other ‘undraped’ model was rendered only in a plain colour, graded using the sequence ‘the higher-the-lighter’ (Fig. 1). A walking route was clearly marked on each map/model, subdivided into five colour-coded sections and depicted by a line located 10m above the model surface to avoid visual confusion with underlying symbols. The ArcScene interface offers interactivity, allowing participants to manipulate the images and view them, either statically or dynamically, from any angle, distance or direction. The extent of zoom is, theoretically, unlimited but is visually affected by the increasing courseness of the raster characteristics of the OS map drape. Zooming in too far gives an unusable fragmentary pixellated image. It is also possible to ‘fly’ across the landscape at any altitude and at various set speeds, but with limited control, using only mouse and keyboard. In other experiments the flat 2D maps would also be viewed with ArcScene but fewer interactive facilities were employed. A Pentium 4 laptop computer was used to deliver the map and model images during the experimental sessions.

To conclude this section it should be pointed out that this is a laboratory-based programme. What is being tested is the use of maps/terrain models in what might be called the planning phase of proposed outdoor journeys. At a later stage it is intended to examine the use of such tools in, for example, PDAs, for the support of on-going wayfinding.

3.2 Experiment 1: Pilot Study

In this experiment only terrain models were used, without comparison with the 2D maps. The focus of interest was on how differing levels of information on the models (the draped and undraped conditions, as described above) would affect their perception, value as an information source and ease of use.

The main hypothesis was that there *would* be a difference between judgements made from each model type. This hypothesis, however, was non-directional, with two possible outcomes:

a. That the version draped with OS data might be easier to use due to the greater level of information (such as grid squares and contours) available to the user and, possibly, greater familiarity by participants with the map-like image.
b. That, if the draped information is ignored (i.e. not used in calculations), estimations would inevitably end up being based on perception of the spatial surface model alone. In fact if this strategy was adopted the draped versions of the models might also overload the user with unnecessary or undesired symbolism and even hinder judgement.

Twenty-four student undergraduates took part (12 female and 12 male), aged between 19 and 31. The experiment was designed with both between- and within-subject comparisons to avoid bias. The tasks (dependent variables) were a) the estimations of the slope angles/gradients of each (coloured) section of the depicted routes, and b) the walking time estimates for each entire route. The independent variable was the model type, draped and undraped.

3.3 Procedure

Following an introductory session which included familiarisation with the computer (using a training map), the participants then completed twelve trials, each with a different model. They were shown each of the twelve landscapes once, six with the ‘draped’ version and six with the undraped.

However map types presented were different for different participants i.e. while half saw models 1-6 draped and 7-12 undraped, the other half received the opposite combination. Also to minimise practice effects, the order in which they were presented was counter-balanced.

In each trial participants were asked to do the following:

1. Observe the walking route depicted on the model, and for each of the colour-coded sections (A-E) were asked to
    a. assign a gradient rating based on a 5-point scale (essentially flat, gentle, medium, steep, extreme)
    b. assign an angle estimation, selecting from a 90^{o} diagram with eight intermediate 10^{o} steps

    The angle judgement was included to allow for personal bias on slope severity *(a 'moderate' gradient for one might be 'steep' for another).* Specific angle estimations could also be compared later with the actual gradients, measured from the maps.
2. Consider the entire route, and estimate the time taken, in hours and minutes, to walk it.

When carrying out these tasks, participants could make full use of the interactive facilities of zooming, rotating, etc., with no time restrictions.

When the experiment session ended each participant completed a short questionnaire to identify any previous experience in map reading or navigation, and also to check for their familiarity with any of the Scottish regions depicted in the models. This was necessary to ensure that judgements had not been made from previous personal knowledge rather than direct observation of the models alone. They were also asked to select which of the two models they had preferred or found easier to use. Although there were no time restrictions during the trials, the overall time taken was recorded for each participant as it could be a factor influencing performance. Estimation data was also retained for later analysis.

3.4 Results

Even after careful analysis of the data, no significant differences in walking times or gradient/angle estimations were detected between the draped and undraped models.

Estimations of gradient also compared well with actual gradient values (computed from the maps) and walking times (from Naismith's Rule[4]). This would show if one model type affected accuracy judgements, but there was still no significant difference in walking time estimations between the two model types. Mean real gradients/angles also correlated well with estimations in both conditions. Times taken by participants when using their 'preferred' **draped** maps were slightly longer, but with no differences in performance. Equally, there were no significant differences in performance between those with high and low self-reported measures of hill-walking frequency in either draped or undraped condition. There was also wide and variable experience/ability in map-reading among the participants, but this was not formally assessed in the first experiment.

3.5 Discussion

The main experimental hypothesis was not supported. No differences were found between responses for the two models, draped and undraped, estimations of gradients/angles being very similar in accuracy and magnitude. These results are striking as they might not have been anticipated! Why were there no differences in response between the model types? The OS data seemed neither to help nor hinder the estimation process. Was this extra information regarded as unnecessary or did it even lead to perceptual or cognitive overload for the user? This is probably unlikely. If it had been a hindrance, greater success would have been recorded with the undraped model, and this was not the case. Also more than 66% of the participants *preferred* the draped model and so this extra information could not have been a problem. Apart from a very few participants who used the OS grid squares to help with distance estimation, most did not use the OS data, as the accuracy ratings were no better in the draped than the undraped condition. Most participants just seemed to estimate. Although the draped models did have more altitude information (contours) this was not used when estimating slope angles. Only if the task had required calculations to be made would the contour data have been of direct use.

Although originally intended merely to test the experimental procedures for the wider programme, this introductory experiment has also raised

[4] Naismith's Rule provides a means of estimating route times in the hills, by taking into account both the distance to be walked and height to be climbed: Time, in minutes *(excluding rest breaks)* = 12 x the distance walked *(km)* + 0.1 x the height climbed *(m).*

some interesting questions which have been carried forward into the main investigations, especially the more formal determination of expertise groups.

4 Conclusion

This introductory study set the scene for the main programme of experimentation which has now been completed and is being reported elsewhere. Although some research studies have been done on the value of animated and interactive terrain modelling in support of various tasks, most have not followed a strictly experimental approach. In view of the rapidly growing popularity of leisure walking (see walkingworld.com) and growing anecdotal evidence of the attractiveness and popularity of new animated and interactive cartographic products (e.g. Memory-Map, Anquet Maps), the authors of this paper advise that the nature and use of the latter should be examined with some scientific rigour. Only through such methods can a suitably rich and dependable source of empirical evidence be assembled which can be of value to:

1. Product designers seeking to assess the degree of realism or symbolic support required from the displays
2. Psychologists seeking to extend their knowledge of visuo-spatial cognition and mental imagery
3. Those responsible for giving advice and training to different categories of mountain visitor
4. Environmental managers and planners.

This could also lead to new and valuable Internet resources and further opportunities to offer advice and pass on information about techniques for planning leisure journeys and for use with PDAs (or new-generation mobile phones!) in the field. It is also hoped that when completed these experiments will contribute to a greater understanding of geovisualization methods in general.

Acknowledgements

This project was supported by funding from the British Academy. Sincere thanks are due to Alan Tinto (School of Psychology, University of Aberdeen) who was the research assistant on the project, and, especially to

David Miller (Macaulay Land Use Research Institute, Aberdeen) who has been a valuable member of the research team.

References

Donald, M., (1991). Origins of the Modern Mind. Cambridge, Massachusetts, Harvard University Press.

Eccles, D.W., Walsh, S.E., & Ingledew, D.K. (2002), A grounded theory of expert cognition in orienteering. Journal of Sport Exercise Psychology, 24(1), 68-88.

Hafting, T., Fyhn, M., Molden, S., Moser, M-B and Moser, E.I., (2005), Microstructure of a spatial map in the entorhinal cortex. Nature, 436, 801-806.

MacEachren, A. & Kraak, M-J. (2001), Research challenges in Geovisualisation. Cartography and Geographic Information Science, 28(1)

Moeser, S., (1988), Cognitive mapping in a complex building. Environment and Behavour, 20, 21-49.

Moore, K., Dykes, J. Wood, J., (1999). Using Java to interact with geo-referenced VRML within a VFC, Computers & Geosciences, 25, 1125-1136.

Morrison, K.W. & Purves, R.S., (2001). Interactive, dynamic landscape visualisation over the web: an illustration utilising Java servlets and VRML. In D.B. Kidner and G.Higgs (Eds.) Proceedings of 9th GISRUK Annual Conference, University of Glamorgan, April 2001, 202-206.

Ottosson, T, (1996). Cognition in orienteering: Theoretical perspectives and methods of study. Scientific Journal of Orienteering, 12, 66-72.

Scaife, M. & Rogers, Y. (1996). External cognition: how do graphical representations work? Internatiional Journal of Human-Computer Studies, 45, 185-213.

Slocum, T., et al, (2001), Cognitive and usability issues in Geovisualisation. Cartography and Geographic Information Science, 28(1).

Thorndyke, P.W. & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. Cognitive Psychology, 14(4), 560-589.

Wood, M., (2003). The mountain panorama and its significance in the Scottish context, Cartographica 38 (1 & 2), 103-118.

Wood, M. (2000). Professor Heinrich C. Berann and his Methods. Proceedings of the Seminar, Mountain Cartography 2000, M. Buchroitner (Ed.) Dresden University, 197-205.

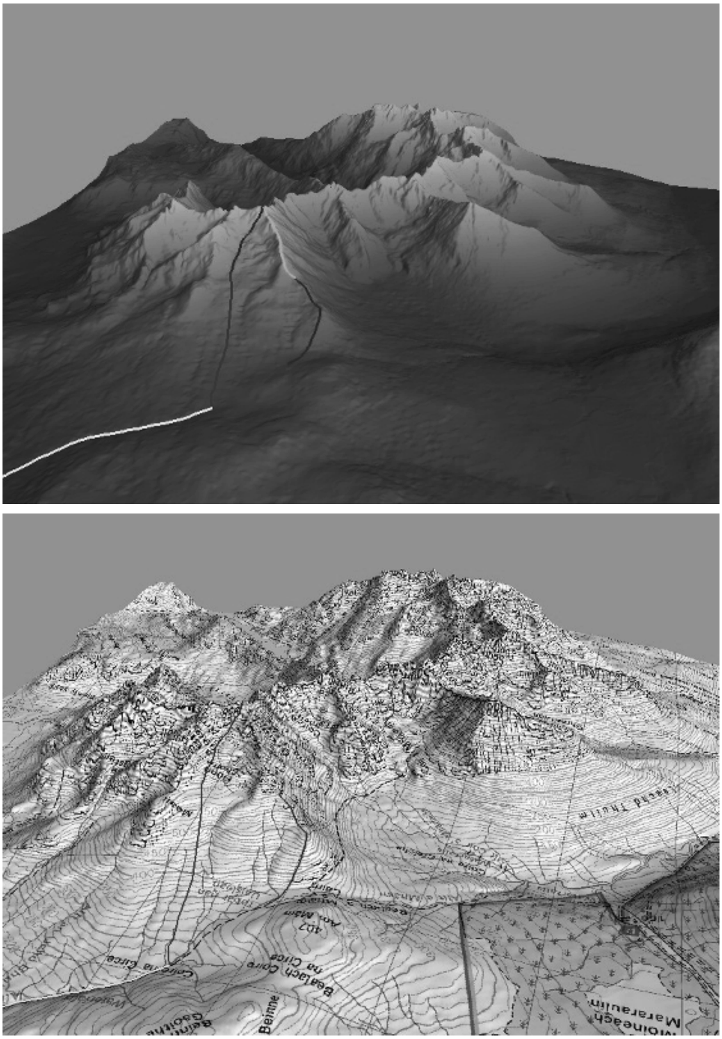
Wood, M. & McCrorie, I. (1993), Panorama maps: an old magic on a new digital plane. Proceedings of the 16th International cartographic Conference, Koln, Germany: 103-113

Wood, M. & Goodwin, A. Harvey, (1995). Digital terrain models in tourism: an investigation into the development and use of computer-generated 3Dviews for route-finding in mountain areas. Proceedings of the 17th International Conference on Cartography, Barcelona, Catalunya.

Zimmer, H.D. (2001). How to optimise human wayfinding: a cognitive science approach to a user-adaptive design. Cognitive Processing, 2-3, 309-322.

URL1: Anquet Maps, www.anquet.co.uk
URL2: Memory-map, www.memory-map.co.uk

Copyright



An Approach towards Semantics-Based Navigation in 3D City Models on Mobile Devices

Jürgen Döllner, Benjamin Hagedorn, Steffen Schmidt

University of Potsdam – Hasso-Plattner-Institut, Potsdam, Germany
{juergen.doellner, benjamin.hagedorn, steffen.schmidt}@hpi.uni-potsdam.de

Abstract

This paper outlines a novel approach for user navigation in complex virtual 3D city models on mobile devices. Users navigate within the virtual 3D city model by sketching navigation commands in the perspective view on the mobile client. The sketches are sent to the server, which reprojects the sketches onto the 3D scene, interprets these sketches in terms of navigation commands, and sends the resulting video-encoded image stream to the mobile client. This approach allows us to provide interactivity for complex virtual 3D city models on resource and bandwidth limited mobile clients. A high degree of usability is achieved because users can trigger complex navigation commands in a task and goal oriented way taking advantage of the navigation properties and affordances inherent to elements of geovirtual environments.

1 Motivation

Virtual 3D city models represent urban spatial and geo-referenced data by 3D geovirtual environments (GeoVE) that include terrain models, building models, vegetation models as well as models of roads and transportation systems. In general, these models serve to present, explore, analyze, and manage these urban information spaces and, therefore, constitute a major user-interface paradigm for 3D geoinformation systems.

An increasing number of applications and systems incorporate virtual 3D city models as essential system components such as for facility management, logistics, security, telecommunication, disaster management, location-based services, real estate portals as well as entertainment and education products. Consequently, a large number of potential users and usages require an efficient and effective mobile access to virtual 3D city models and their contents.

We present a novel solution for accessing virtual 3D city models on mobile devices. The user controls the navigation within the virtual 3D city model by navigation command sketches drawn directly on the view-plane of the mobile client (Fig. 1). The sketches are sent to the server, which reprojects the sketches onto the 3D scene correlating the sketches to scene objects, interprets these sketches in terms of navigation commands, and sends the resulting video-encoded image stream to the mobile client. That is, the mobile client enables users to specify and retrieve step-by-step created video sequences that correspond to their navigation intentions.

Fig. 1. Sketching the navigation command “look around” (left). Sketching the navigation command “walk along the path and, finally, look at the indicated building” (right).

2 Related Work and Challenges of Mobile 3D City Models

Mobile applications of virtual 3D city models represent a major and complex research challenge due to limited bandwidth and graphics capabilities, restricted interaction capabilities, data standardizations and distribution techniques, and digital rights issues.

2.1 Mobile 3D Rendering

In 3D computer graphics, numerous rendering techniques are available to cope with complex virtual environments, including discrete and continuous multi-resolution geometry and texture representations, view-frustum culling, occlusion culling, imposter techniques, and scene-graph optimizations (Akine-Möller and Haines 2002). Virtual 3D city model visualizations require an efficient management of large-scale texture data, e.g., for aerial photography and building facades (Buchholz and Döllner 2005), and level-of-detail management for large heterogeneous 3D object collections (Davis et al. 1999) and 3D terrain surfaces (Döllner et al. 2000). Although these rendering techniques enable real-time rendering of complex 3D scenes, they generally cannot be transferred directly on mobile devices due to limited computational resources and power.

One principal approach to efficient *mobile 3D rendering* consists in the adaptive, progressive, and compressed transmission of 3D graphics data to mobile clients. For example, Royan et al. (2003) describe client-server architecture for mobile 3D virtual city visualizations based on a progressive and hierarchical representation for GeoVEs. The server pre-computes multi-resolution representations of terrain models and building models, and progressively sends data about visible areas to the mobile clients. The clients allow users to interact with the 3D city model (e.g., virtual walk-throughs, fly-overs, etc.). However, the limited 3D graphics acceleration on today’s mobile devices makes it difficult to implement fully featured 3D rendering techniques for virtual 3D city models. Furthermore, the implementation is complicated due to the broad variety of hardware and software solutions for mobile 3D graphics (e.g., OpenGL ES, Mobile 3D Graphics API for J2ME).

Another principle solution consists in *server-side 3D rendering* and the progressive, compressed transmission of image sequences. For example, Cheng et al. (2004) investigate a client-server approach for visualizing complex 3D models on thin clients applying real-time MPEG-4 streaming to compress, transmit, and visualize rendered image sequences. They identify the MPEG-4 encoding speed as bottleneck of client-server 3D rendering, and devise a fast motion estimation process for the MPEG-4 encoding process.

2.2 Mobile 3D Interaction

To achieve a high degree of usability, mobile applications require goal-oriented and task-oriented interaction techniques that take into account the specific restrictions of mobile devices, e.g., no mouse, no desk, or one-handed operation. For this reason, approaches for automating user interaction are crucial for effective mobile user interaction. Of course, these approaches are also faced with the general problems of navigating in virtual worlds (Russo et al. 2000).

A critical task in applications of GeoVE represents the process of navigation, “whereby people determine where they are, where everything else is, and how to get to particular objects or places” (Jul and Furnas 1997). Navigation as the primary interaction can be distinguished into three kinds, naive search, targeted search, and exploration (Darken and Sibert 1996) and serves to explore, analyze, and gather geoinformation as well as to trigger object-specific interaction. To do this, users generally move the virtual camera or an avatar through the GeoVE. This way the user builds up a mental model of the GeoVE by forming linear maps and combining them to spatial maps (Ingram and Benford 1995). Wernert and Hanson (1999) incorporate task-based constraints on the navigation parameters (e.g., viewer position and orientation) to enable the designer of GeoVE “to provide extra assistance to keep the user’s explorational wanderings and attention focused on the task objectives”.

Common navigation controls for GeoVE include world-in-hand controls, fly-over controls, and virtual trackballs. Burtnyk et al. (2002) introduce a general approach of facilitating navigation in GeoVE based on explicitly designed navigation spaces using integrated spatial and temporal controls. Buchholz et al. (2005) describe a concept of smart and physically-based navigation techniques, controlling the user’s movement similar to an assistance system preserving users from being disoriented or getting lost in the GeoVE. It constrains the movements to be inside the GeoVE, hinders collisions with buildings, controls the gaze direction at the terrain borders, and facilitates the switch between navigation modes. For mobile applications, semantics-based navigation control can integrate and extend these concepts.

Igarashi et al. (1998) develop an intuitive approach for specifying navigation commands: The user draws the intended navigation path as a curve on the view plane. This path is mapped to the 3D scene and determines the 3D path the avatar moves along. This way, the user can specify not only the final position, but also the route and the camera direction at the goal with a single stroke. Our approach also has been motivated by the metaphor-aware 3D navigation technique (Russo et al. 2000) and specialized for virtual 3D city models.

2.3 Standardization and Distribution

Applications of virtual 3D city models also suffer from a lack of data standards and flexible distribution techniques. Virtual 3D city models frequently are implemented as graphical models without explicitly modeling semantic and topological relations. Therefore, they can almost only be used for visualization purposes but not as a data basis for higher-level functionality such as simulations, analysis tasks, or spatial data mining. The limited reusability and interoperability inhibits the broader use of virtual 3D city models. CityGML represents a first XML and GML-based format for storing and exchanging virtual 3D city models (Kolbe et al. 2005), which also represents semantic and thematic properties, taxonomies and aggregations.

With respect to distribution, a complete delivery of city model data would result in massive data transfers. Even if only a part of a complex virtual 3D city model is required (e.g., view-dependent multiresolution selections), the costs for geometry and texture data for high-quality photorealistic models typically exceed current and future transmission capabilities.

2.4 Digital Rights Management

Protecting the contents of 3D city models is one of the most critical aspects of real-world business models underlying 3D city model applications (Döllner 2005). The transmission of raw city model data or derived detailed 3D graphics data imply severe drawbacks for copyright issues and controlling usage and distribution. For this reason, we transmit only video sequences but no raw data to the mobile clients.

3 Sketch-Based Navigation Commands

3.1 Real-Time Interaction vs. Selective Interaction

Common navigation techniques allow users to control their movement within the virtual environment in real-time. For mobile devices, however, real-time 3D rendering of virtual 3D city models is not practically possible due to limited computation resources and bandwidth as well as the non-stable data transmission. In contrast to the real-time user reaction that characterizes most games taking place in virtual environments, real-time 3D interaction is not crucial for many applications and systems of virtual

3D city models because exploration and analysis tasks performed by users are based on a selective, targeted access of spatial information. That is, the delay between issuing interaction commands on the mobile device and the execution of the commands is acceptable and corresponds to the expectation of the user.

3.2 Concepts of Sketch-Based Navigation

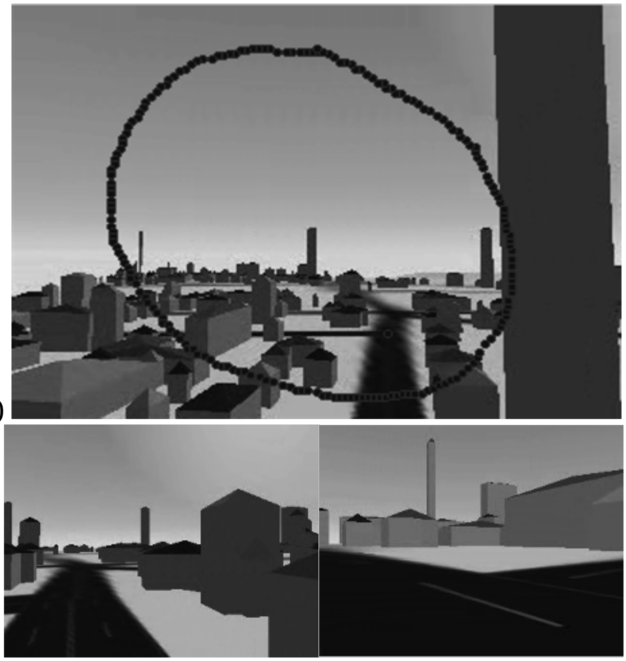
In our approach, navigation commands are graphically specified in the perspective view of the virtual 3D city model on the mobile client, e.g., drawing a line along a street, pointing to a building or the sky. The sketches are correlated with the objects of the GeoVE by reprojecting the sketched shapes onto the 3D scene. The sketch-based navigation commands are interpreted based on the semantics of sketch-correlated objects and their inherent navigation affordances.

We distinguish between three types of information used for interpreting the commands:

- *Spatial context*: The spatial context refers to the virtual location to which the sketch is aligned or associated. For example, the user can draw a path along a street or mark a specific building.
- *Temporal context*: The temporal context refers to the order in which the user composes the sketch elements. For example, as first step the user draws a path along a street, and then marks the building.
- *Sketch geometry*: The elements include points, lines, and polygons drawn in the perspective view. They can be grouped and interpreted by higher-level sketch geometry such as circle-like paths or u-like paths.

The sketches can be differentiated into *location-aware sketches* and *gestures*. Location-aware sketches refer to a spatial context, whereas gestures do not. From a technical point of view, gesture recognition requires large, screen-wide drawings for correct identification. For example, a circle gesture cannot be drawn close to the corners of the screen. Gestures are known from computer games, from several navigation-aware applications, or as utility programs that can be used for desktop interaction.

We allow for concatenating and building up a temporal context for the navigation command sketches. For high usability and consistency of the user interface it must be considered that sketches might a) represent a place to go to or a path to go along, b) mean one ore more points to gaze at, or c) conclude both, place and direction of view. By combining gestures with other sketches we can introduce spatial context to gestures (Fig. 2).



a) b) c)

Fig. 2. The user sketches a point on a junction and adds a circle-like gesture as secondary input (a). In the resulting animation, the camera sinks down to the marked junction (b) and then performs a turn-around (c).

3.3 Sketch-Based Navigation Vocabulary

A first collection of spatial and temporal contexts together with sketch elements is illustrated in Table 1. Gabbard (1997) points out that “when assessing metaphors for navigation and locomotion in VEs, it is important to consider mappings of integral navigation and locomotion components to metaphor gestures or mechanisms.” The sketch-based navigation commands within their spatial and temporal context provide such mechanisms. For example, drawing a single, straight path along a street object indicates, “walk along the street”. A circle-like (close or nearly closed) path drawn on the terrain surface indicates “look around” using a drawn point as camera position. A path drawn along a street with a final indicated u-turn indicates, “walk along to the end of the street, turn around, and walk back”.

Table 1. Overview of sketch-based navigation commands

Name & Context	Example Sketch	Navigation Sketch	Navigation Action
Point-House		Point on a building.	Finding shortest path to the building, going there, and looking at the building.
Point-Roof		Point on a building's roof.	Flying up to the roof, placing the camera on top, and looking around.
Curve-Street		Curve or polygon on a street.	Walking on street and looking back finally.
Curve-Street, Point-House		Curve on a street and point on a building.	Walking along the street and looking at the building finally.
Point-Ground, Point-House		Point on the ground and point on a building.	Flying to the marked ground point and looking at the building finally.
Point-Street, Circle Gesture		Point on the street and circle-shaped gesture.	Flying to the marked ground point and looking around.

Point-Sky		Point on the sky.	Soaring above ground for overview.

4 Client-Server Architecture for 3D Visualization

The presented approach has been implemented based on a client-server architecture outlined in Fig. 3. We assume that the 3D city model is hosted on the server and that the mobile devices efficiently support encoded video streams.

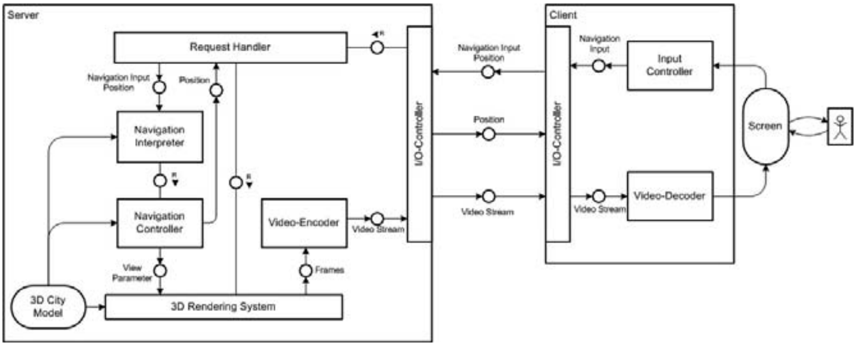


Fig. 3. Client-server architecture of our system for mobile access to virtual 3D city models

4.1 Server System

The server system is responsible for handling requests sent from the clients, for interpreting and controlling sketch-based navigation commands, for 3D rendering, and the video-stream encoding. It provides a web service interface to the virtual 3D city model. The clients can communicate with the server by exchanging SOAP messages. The interface supports three main operations:

- *GetCapabilities*: Provides the service metadata including information about the used streaming protocol and available start positions. The clients call the operation at the beginning of the communication.

- *GetStartPosition:* Renders and transmits an image of the start position. This image provides the spatial context for the user's first navigation inputs.
- *GetMotion:* Interprets sketch-based navigation commands and initiates the rendering of the camera animation. Because the server is stateless, it has to reconstruct what the user of the client saw while drawing the sketches. Therefore, the request contains the final camera position of the preceding request. If the user stopped before the end of the animation for a new input, the client's position can be determined by the start position and inputs of the previous navigation and the point in time the user stopped the video. Both camera positions, at the beginning and the end of navigation, are included in the response message.

The *navigation interpreter* detects the semantics of sketch-correlated scene elements and identifies the classes to which hit objects belong. For a sketch that has more than one input point we determine which object type occurs most frequently such as in the case of a path on a street whereby not all of the input points are placed exactly. The *navigation controller* calculates the resulting animations (Christianson et al. 1996; Mackinlay et al. 1990). Special navigation controllers use a navigation network geometry that provides paths that can be used to walk along. The position to look at a house is determined as the nearest point on such a path element. For an effective 3D overview for the rise-to-the-sky navigation we provide a map of view directions as introduced by Hanson and Wernert (1997). It allows us to determine a suitable direction to look at from a specific point to gain as much spatial information as possible. The current implementation is based on height-defined landmarks. The *rendering component* encodes the animation frames into an MPEG-4 video stream (Cheng et al. 2004; Noimark and Cohen-Or 2003) by the *video encoder*. The resulting video stream is transmitted to the client immediately using a standard streaming protocol. So, the client can start the video playback as soon as possible.

We have implemented and tested a server that uses DIME attachments to deliver the video to the client. The DIME standard is similar to MIME and defines a way to send arbitrary binary data along with SOAP messages. Because the data is sent in chunked data blocks, it allows starting the streaming of the produced video while the rendering process has not been completed. Instead of DIME, any other streaming protocol could be used. In this case, the server's response message must include the parameters necessary to connect to the streaming protocol or server.

4.2 Client System

The thin client system does not contain application logic, it only needs capabilities for receiving and playing the MPEG-4 video streams, capturing the user input and sending and receiving SOAP messages. While receiving a video stream, the client simultaneously decodes and displays the video. Most mobile devices provide built-in support for these tasks. For drawing new navigation sketches, the user can wait for the end of the video or he can stop it at any point of time.

The client records the user inputs as a set of 2D points representing the screen coordinates of the navigation sketches. The temporal context of the input can be determined by the drawing order of the points and the classification of single sketches.

5 Conclusions

We have presented an approach towards semantics-based navigation control for mobile virtual 3D city models. A high degree of usability is achieved because sketch-based navigation commands allow users to trigger effectively complex navigation intentions taking advantage of the navigation properties and affordances inherent to elements of geovirtual environments. In addition, it is perfectly suited for the input devices and usage situation of mobile devices where generally no mouse and no desktop can be assumed. From a technical perspective, the presented approach allows mobile applications to provide users interactive access to complex 3D city models including high-resolution 3D terrain geometry, 3D building geometry, and textures exceeding several hundreds of GB of storage. In particular, the server can be optimized for processing large-scale 3D city models using high-end computer graphics hardware, whereas only multimedia capabilities are required from the mobile client.

In our future work, we will address general sketch-based interaction commands and visual feedbacks about the automated navigation. We also plan to include in the video stream meta-information about the scene and its objects.

References

T. Akenine-Möller, E. Haines, “Real-Time Rendering”, 2^{nd} Ed., A K Peters, 2002.

H. Buchholz, J. Bohnet, J. Döllner, “Smart and Physically-Based Navigation in 3D Geovirtual Environments”, IEEE Inform. Visualisation, London, 2005.

H. Buchholz and J. Döllner, “View-Dependent Rendering of Multiresolution Texture Atlases”, Proc. IEEE Visualization, IEEE CS Press, 2005, to appear.

N. Burtnyk, A. Khan, G. Fitzmaurice, R. Balakrishnan, and G. Kurtenbach, “StyleCam: Interactive Stylized 3D Navigation using Integrated Spatial & Temporal Controls”, Proceedings of the ACM Symposium on User Interface Software and Technology, 2002, pp. 101-110.

L. Cheng, A. Bhushan, R. Pajarola, and M. El Zarki, “Real-Time 3D Graphics Streaming using MPEG-4”, BroadWise ’04, San Jose, CA, USA, July 2004.

D.B. Christianson, S.E. Anderson, L. He, D.H. Salesin, D.S. Weld, and M.F. Cohen, “Declarative Camera Control for Automatic Cinematography”, Proceedings of AAAI '96 (Portland, OR), 1996, pp. 148-155.

R.P. Darken and J.L. Sibert, ”Wayfinding Strategies and Behaviors in Large Virtual Worlds”, Proceedings of ACM SIGCHI 96, 1996, pp. 142-149.

D. Davis, W. Ribarsky, T.Y. Jiang, N. Faust, and S. Ho, “Real-Time Visualization of Scalably Large Collections of Heterogeneous Objects”, Proceedings of IEEE Visualization 1999, pp. 437-440.

J. Döllner, “Constraints as Means of Controlling Usage of Geovirtual Environments”, Cartography and GIS, 2005, 32(2):69-79.

J. Döllner, K. Baumann, and K. Hinrichs, “Texturing Techniques for Terrain Visualization”, Proc. IEEE Visualization, IEEE CS Press, 2000, pp. 207-234.

J.L. Gabbard, “A Taxonomy of Usability Characteristics in Virtual Environments”, Technical Report, Virginia Polytech. Institute, 1997.

A.J. Hanson and E.A. Wernert, “Constrained 3D Navigation with 2D Controllers”, IEEE, Proceedings on the Conference on Visualization, 1997, pp. 175-183.

T. Igarashi, R. Kadobayashi, K. Mase, and H. Tanaka, “Path Drawing for 3D Walkthrough”, 11th Annual Symposium on User Interface Software and Technology, ACM UIST'98, San Francisco, November 1998, pp.173-174.

R. Ingram and S. Benford, ”Legability Enhancement for Information Visualization”, Proceedings of IEEE Visualization 1995, pp. 209-216.

S. Jul and G.W. Furnas, “Navigation in Electronic Worlds: A CHI 97 Workshop”, ACM SIGCHI Bulletin, Vol. 29, No. 4, October 1997, pp. 44-49.

T.H. Kolbe, G. Gröger, and L. Plümer, “CityGML - Interoperable Access to 3D City Models”, Internat. Symp. on Geoinfo. for Disaster Management, 2005.

J.D. Mackinlay, S.K. Card, and G.G. Robertson, “Rapid Controlled Movement Through a Virtual 3D Workspace”, Proceedings of SIGGRAPH '90, August 1990, 24(4):171-176.

Y. Noimark and D. Cohen-Or, “Streaming Scenes to MPEG-4 Video-enabled Devices”, IEEE Computer Graphics and Applications, 2003, 23(1):58-64.

J. Royan, C. Bouville, and P. Gioia, “PBTree - A New Progressive and Hierarchical Representation for Network-Based Navigation in Urban Environments”, Proc. Vision Modeling Visualization, pp. 299-307.

C. Russo dos Santos, P. Gros, P. Abel, D. Loisel, N. Trichaud, and J.-P. Paris, “Metaphor-Aware 3D Navigation”, InfoVis'2000, IEEE Symposium on Information Visualization, Salt Lake City, October 2000, p. 155.

E.A. Wernert and A.J. Hanson, “A Framework for Assisted Exploration with Collaboration”, Proc. IEEE Visualization, 1999, pp. 241-248.

The World as a User Interface: Augmented Reality for Ubiquitous Computing

Dieter Schmalstieg[1], Gerhard Reitmayr[2]

[1] Graz University of Technology, schmalstieg@icg.tu-graz.ac.at
[2] University of Cambridge, gr281@cam.ac.uk

Abstract

We discuss the possibilities of augmented reality (AR) as a ubiquitous user interface to the real world. A mobile AR system can constantly provide guidance to its user through visual annotation of the physical environment. The first part of the paper discusses the necessary ingredients for ubiquitous AR, on which we have worked in the recent past, namely mobile AR hardware, wide area tracking, unobtrusive user interfaces, application prototypes, and geographic data models suitable for AR. The second part of the paper examines future requirements of such data models in greater detail. Based on the lessons learned in our previous work, we identify shortcoming of existing standards for geographic information systems and visualization models. Ubiquitous AR requires independence of the data model from specific applications and their implicit assumptions. A semantic network model of geo-referenced data provides such a data model. We examine how such a model fits the requirements of AR applications, and how it can be implemented in practice.

1 Introduction

Augmented reality (AR) is an excellent user interface for mobile computing applications, because it allows intuitive information browsing of location-referenced information. In an AR environment, the user's perception of the real world is enhanced by computer-generated entities such as 3D objects and spatialized audio [Azu97]. Interaction with these entities occurs in real-time providing convincing feedback to the user and giving the impression of natural interaction. Augmented reality as a user interface becomes particularly powerful when the computer has access to location-based information so that it can seemingly merge virtual entities with real world objects in a convincing manner.

Over the last years, we have created a mobile augmented reality system and a set of applications to gather experience of ubiquitous augmented reality applications. We focused on navigation and created solutions for both indoor and outdoor navigation. Both applications require extensive 3D models and information which is presented to the user. Accurate and complete models of buildings and their interiors are required for rendering occlusions and highlights of buildings. Different navigation models for indoor or outdoor use were developed to fit specific requirements.

The first part of this paper summarizes our work on mobile indoor and outdoor AR: presenting various hardware platforms for mobile computing and graphics; hybrid indoor/outdoor tracking solutions; user interface considerations and test applications. The second part introduces the concept of a semantic world model for AR, which directly derives from an analysis of the requirements of applying AR techniques within ubiquitous computing applications. We review recent developments in the geographic information systems (GIS) community, and how they can be used by mobile AR systems. As part of this approach, we present a conceptual design for a semantic network for AR, which can serve as a computational back end providing a new level of contextual information for AR and other types of ubiquitous computing services.

Fig. 1. The backpack system is based on a conventional notebook computer displaying stereoscopic graphics overlays in an optical see-through head mounted display.

Fig. 2. This handheld AR platform is based on a mini tablet PC that is operated as a video-see through "magic lens" device. It currently supports five different tracking technologies (3 shown) and weighs less than 1.5kg.

Fig. 3. The smallest fully featured AR platform to date is based on commercial PocketPC 2003 handheld computers. Using an embedded camera, these devices can perform optical marker tracking and real-time 3D graphics at ~20fps. Shown is the “Invisible Train” multi-user game successfully demonstrated at SIGGRAPH 2004.

2 Mobile augmented reality platform

2.1 Hardware and software

The successful delivery of mobile Augmented Reality (AR) is an ongoing challenge, as interactive 3D applications must be implemented on constrained hardware platforms, requiring tracking over a large area of operation at high accuracy. While earlier work on mobile AR uses backpack prototypes, e.g. [FMH*97] or [TDP*98], more recently there has been a trend towards smaller, discreet, lightweight handheld setups based on PDAs and cell phones. We have assembled a series of experimental platforms of varying form factor and capability for real-world testing of the trade-offs in wearable AR.

Our implementation of a classic backpack solution (Fig. 1) involved a frame equipped with a notebook computer running a standard operating system (specifically a Dell Precision 8100, 2GHz P4 CPU, NVidia Quadro4Go, Windows XP). This standard platform provides standard interfaces for communication (802.11 wireless LAN, GPRS) and peripherals. A differential GPS receiver (Trimble Pathfinder Pocket or Garmin GPSMouse) is used to determine the position of the system in outdoor applications. We use an optical-see-through stereoscopic color head mounted display (Sony Glasstron D100-BE) fixed to a helmet as an output device. The helmet also supports an inertial orientation tracker (InterSense InertiaCube2) and a camera (PointGrey Firefly) for indoor fiducial tracking and video see-through configurations [RS04].

While the backpack system uses very powerful hardware and allows development directly on the target platform, its weight and ergonomic properties are clearly unsatisfactory. We are therefore also experimenting with smaller handheld computer platforms, which allow a “magic lens” style of video see-through augmentation. Such a handheld AR platform is inexpensive and ergonomically superior to the backpack solution. Most potential users are already familiar with camcorders and consequently understand the handling (hand-eye coordination) of a handheld video-see through device. Informally we have observed that users prefer handheld AR over head mounted displays despite the lack of stereoscopic graphics and hands-free operation. The lower computational power of handhelds is partially compensated by the reduction in graphical complexity: monoscopic rather than stereoscopic; smaller screens, increased tolerance of lower resolutions and frame rates.

We have developed two handheld setups. The first setup (Fig.2) is based on a mini tablet PC (Sony Vaio U70, 1GHz Pentium-M, Windows XP). This platform combines a regular PC compatible computer with several peripherals into a very compact form factor (footprint 15x20cm, 1400g including peripherals). The second setup (Fig.3) is based on the Pocket PC standard for personal digital assistant (PDA) computers (ARM9 CPU currently attaining maximum speeds of 624MHz, Windows CE). While these devices weigh only around the 180g mark, they still feature a touch screen and a built-in camera. We have managed to implement real-time optical tracking and 3D graphics on the Pocket PC platform [WS03]. In terms of weight, size, and price these devices are almost ideally suited for our purposes, but software development for PDA operating systems is still not a straightforward task.

The software framework that allows rapid prototyping of AR applications with a high degree of 3D interaction is *Studierstube*, a versatile environment for developing virtual reality and AR applications [SFH*02]. It is based on an object-oriented scene graph (Coin), which allows the description of 3D scenes and 3D interactions through convenient declarative scripting.

2.2 User Interface

The main use of a mobile AR device is as an information appliance operated in browsing mode. The system should provide context-sensitive cues while the user is busy performing a task or navigating through the environment. Consequently, most input to the system should be automatically derived from situational context, without requiring explicit user attention. The main method for achieving this is by tracking the user's position in the environment, and the user's current viewing direction. Consequently, wide area tracking is of major importance for a mobile AR system (see next section).

However, applications will still require a certain amount of direct control. For example, a user may want to select a navigation target from a list of addresses or a map, or if the destination is within visible range more directly by using gaze direction. For such explicit control, we have investigated a number of interface alternatives.

Touch pad and touch screen: handheld AR devices are already equipped with a touch screen, which can be conveniently used to display on-screen menus operated by stylus or a finger (Fig. 1). The same touch screen can be used for selecting objects in the video-see through display by tapping on their position on the screen, effectively a form of raypicking interaction. For configurations using a head-mounted display, we have relied on either a handheld touchpad peripheral that is used to control a 2D cursor in the heads-up display, or by an additional PDA which can display menus directly on its screen. Both the touchpad and the PDA can be tracked using fiducial markers observed by the helmet-mounted camera, if 3D input is desired.

While the touch screen interaction clearly hints at its origins in desktop 2D and 3D user interfaces, the iOrb device (Fig.4) was specifically designed for unified command and spatial input for mobile AR [RCK*05]. It consists of a single 3DOF inertial tracker (XSens MT9) embedded in a shell composed of two hemispheres of about 8cm diameter. By turning the sphere and then pressing the two hemispheres together, the user can issue application commands using variants of 1D and 2D pie menu techniques. Similarly, spatial selections can be made using a picking ray or cone. All interactions use only relative rotational measurements and are therefore mostly insensitive to measurement inaccuracies and drifting.

Fig. 4. The iOrb, an interaction device specifically designed for mobile AR users, unifies command input and 3D spatial selection at a distance through raypicking.

2.3 Tracking infrastructure

Wide area tracking cannot generally be done with a single sensor, because no single tracking technology can provide the range and accuracy required by a general mobile AR system. Therefore a significant body of work exists on hybrid tracking systems, combining multiple tracking technologies through sensor fusion techniques. However, most research focuses on building a single, improved tracking system through sensor fusion, rather than on alternating between different sources depending on availability, location, and context. Notable exceptions are the systems built by Hallaway et al. [HHF*04] and by Piekarski et al. [PAT*04], which are the only systems we are aware of that are capable of alternating between indoor and outdoor operation.

Several previous approaches combine multiple sensors popular for AR setups, such as fiducial tracking, inertial and GPS sensors. However, these approaches lack a general approach for management of arbitrary sensors. We have therefore developed a ubiquitous tracking framework [NWB*04], that addresses the problem of tracker integration and arbitrary sensor management. The most recent integrated solution that uses this approach executes on a handheld computer (U70), and combines five tracking systems: Inertial (XSens MT9), infrared vision (ARTTrack), magnetic (Flock of Birds), optical (ARToolKit), and GPS (Garmin GPS18 USB).

Each tracking technology has a dedicated working volume, with the exception of the inertial orientation tracker. The inertial tracker is used to assist other tracking systems with dead reckoning information, in particular the outdoor GPS system, which does not deliver any estimates of orientation.

As the core tracking software, we use OpenTracker (OT), which implements a pipes-and-filters network for connecting producers and consumers of tracking information [RS01]. The nodes of this network can execute on different hosts in a network. In particular, ARTTrack and Flock of Birds are stationary systems with a dedicated device server, each executing an instance of OT. The OT server will then communicate the tracking information to another instance of OT on the handheld computer over a wireless network.

Outdoor tracking: information is provided by a GPS receiver with the XSens inertial tracker providing complementary orientation estimates. Differential GPS corrections are obtained through a wireless internet connection from a local base station service or the new global correction service.

ARTTrack is a commercial multi-camera system capable of tracking target bodies composed of 4-5 small retro-reflective balls. These lightweight target bodies can easily be mounted on a helmet or handheld device, which can then be tracked in an outside-in mode while in view of the cameras. The mobile device itself is completely passive and does not require batteries or tethered cables for tracking. The pose data is transmitted from a stationary tracking server that performs the online pose estimation to the mobile system using wireless networking.

Flock of Birds: Since the Bird is wired, when the user with the handheld enters the effective envelope of the Bird, it is necessary to physically attach the Bird to the object of interest using velcro straps. Currently, the user triggers an event, by pressing a button, to acknowledge the presence of the Bird. Future versions will automatically determine Bird activity using correlated motion from an inertial tracker permanently mounted to the handheld device and the Bird, once it is attached to the handheld device. Similarly to the ARTTrack, the tracking data is transmitted wirelessly from a stationary server to the mobile device.

ARToolKit: A significantly modified version of the popular vision tracking library *ARToolKit* is used for tracking indoor regions beyond the reach of the magnetic and infrared technologies. Pose estimation of the mobile camera rigidly mounted on a helmet or handheld device is performed using the 2D location of the corners of one or several square markers visible in the camera image. The identity of the markers is then decoded directly from self-correcting 2D barcodes in the marker's interior area, which makes it possible to uniquely discriminate a large number of markers dispersed throughout the indoor environment. By looking up the geometric position of each observed marker in a previously surveyed model of the indoor area, the global pose of the mobile device can be computed from the camera pose estimation. The inertial tracker provides orientation dead reckoning if no markers are observed.

Tracker selection: The OT configuration in the client is responsible for making the tracking "ubiquitous", i.e., permitting online selection of the best available tracking technology. The selection mechanism is currently based on priorities: The stationary tracking technologies ARTTrack, Flock of Birds, and ARToolKit are selected in this order if data from the corresponding sensors is available. The systems attempts to fall back to GPS if none of these technologies is available, and finally reverts to a static map when there is no location data whatsoever. An improved version of the tracking framework is currently under development, which will permit fully automatic discovery of new tracking services based on a tracking service characterization.

2.4 Applications

2.4.1 Outdoor applications

The needs and requirements of a tourist are a suitable starting point for testing location-based applications. A tourist typically has little or no knowledge of the environment. However, tourists have a strong interest in their environment and also want to navigate through their surroundings to visit different locations. Guided tours are also popular for tourists. Consequently, we have chosen a tourist guide application for the City of Vienna as an example scenario for an AR application that integrates a large amount of data from different sources.

Fig. 5. An outdoor navigation guide for a pedestrian visualizes the selected route as a series of waypoints. Note the correct occlusion between the route and the archway .

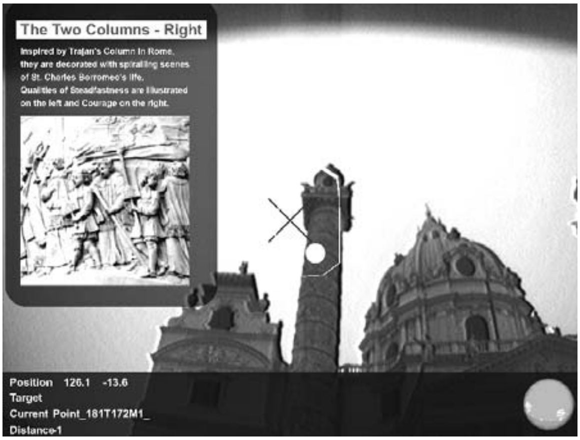


Fig. 6. By gazing at a cultural artefact of interest, a user can instantly recall historic multimedia information.

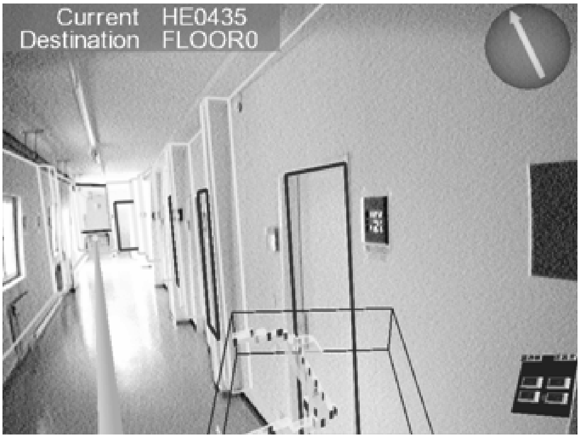


Fig. 7. Indoors, a combination of directional arrows, highlighting of exit, compass (upper right) and world-in-miniature (lower middle) is used to help a user navigate the environment.

Fig. 8. The "Augmented Library" allows a user to quickly locate the shelf on which a book of interest can be found, by emphasizing the relevant location.

The system provides a navigational aid that directs the user to a target location. An information browser displays location-referenced information icons that can be selected to present more detailed information in a variety of formats. Both functions support collaboration between multiple mobile users.

In navigation mode the user selects a specific target address or a desired target location of a certain type such as a supermarket or a pharmacy. The system then computes the shortest path in a known network of possible routes. It is interactive and reacts to the user's movements. It continuously re-computes the shortest path to the target if the user goes astray or decides to take another route.

The information is displayed as a series of waypoints that are visualized as cylinders standing in the environment (Fig.). These cylinders are connected by arrows indicating the direction the user should take between waypoints. Together they become a visible line through the environment that is easy to follow. The user can enable an additional arrow that points directly from her current position to the next waypoint. Buildings can clip the displayed geometry to enable additional depth perception cues between the virtual information and the real world. Finally, the system displays simple directional information, if the user is not able to perceive the next waypoint because she is looking in the wrong direction.

The information browsing mode presents the user with location-based information. Location-referenced information icons appear in view and are selected by looking at them (Fig.). They then present additional information associated with that location. The application conveys historical and cultural information about sights in the City of Vienna.

2.4.2 Indoor applications

Indoors, the mobile AR system guides a user on the way through the long and winding corridors of a university building with direction and map based contextual information [RS03]. The system continuously provides the user with two modes of visual feedback: a heads-up display with directional arrows and a world in miniature model of the environment. The heads-up display shows a wire frame model of the current room superimposed on top of the real scene. The application uses a shortest path search on an adjacency graph of the building to determine the next door/portal on the way to the destination, which is then visually highlighted. In addition, an arrow shows the direction to the next door or portal turn (Fig.). The application also presents a world in miniature model of the building to the user in the lower area of the heads-up display. While the 3D overlay only shows the next exit, the miniature model shows a overview of the user's current environment including the complete path.

In the library, the mobile AR system may assist a user in locating and retrieving a book. Using a menu system, the user can select a book from a database. The corresponding bookshelf is highlighted in the heads-up display to aid the user in finding the book (Fig.). The system can also help in the return of a book. If a marked book is identified, the book's designated shelf is once again highlighted aiding the user to return the book to its correct position. This enables the user to simply look at the book in her hand and trigger the appropriate application behavior.

3 First experiences with large scale data models for AR

The indoor and outdoor applications presented in the last section require extensive 3D models and information presented to the user. Accurate and complete models of the buildings interiors and overall shape are required for rendering occlusions and highlights of buildings. Different navigation models for indoor or outdoor use were developed to fit the specific requirements.

Both applications are supported by a common world model based on an XML description of the geometry of world features. The XML tree is interpreted in the standard geometrical way, by defining a child's pose relative to its parent. However, the open XML-based format is not bound to any particular visualization tool or platform, and affords the definition of other than spatial relations by using relational techniques such as referring to object ids and annotations.

An outdoor AR system can be considered as a special case of a geographic information system (GIS). It presents geo-referenced information in real-time and in 3D, based on the physical location of the user, user preferences, and other context-dependent information. Large amounts of geo-referenced information, such as a 3D world model, require a database system for efficient storage and retrieval. The introduction of a GIS database also solves the problem of providing a consistent view of the 3D world model for a potentially large number of wirelessly connected clients.

Depending on location and context, only a small subset of the information contained in the GIS database is necessary for the client. Information is therefore retrieved dynamically by querying a database server. The response to such a GIS database query typically undergoes a series of transformation steps. Common operations are filtering according to geographic and logical constraints (e.g., return all coffee shops in a radius of 100m), and translation from a more generic format to a data structure that can be directly visualized [VZ04].

When we started with the implementation of our mobile AR framework, solutions for 3D GIS visualization were not sufficiently advanced for our purposes. Consequently, we developed our own XML based data format and processing pipeline using XSLT for data translation. Recently, the emergent Geographic Markup Language version 3 (GML3) standard together with the Web Feature Service (WFS) standard provide a standardized and extensible interface for accessing GIS information. CityGML [KG03], an application profile (extension) to GML3, is similar in spirit to our own custom XML dialect. Since it can be expected that these new standards will be widely supported in the near future, it is advisable to adopt them for AR applications as well.

4 Automating visualization generation

The complexity of interactive visualizations demands automated methods for generating engaging presentations. The fundamental idea of automated visualization generation for AR is quite old [FMS93], but few or no tools for this purpose exist, and GIS technology does not directly resolve this problem either.

Consider, for example, recent research which focuses on specialized visualization techniques for augmented reality, such as communicating the distance of occluded objects [FAD02], improving the readability of text overlays [LT04], providing automated layout of presentation items [HFH*01], filtering information [JLB*00], or adapting the visualization in the presence of tracking errors [CMJ04]. All these techniques could be made generally available to AR applications by an automated approach for visualization generation. However, a simple XML-based world model as presented in the last section is not flexible enough for this purpose.

GML's approach of an explicit representation of the geometric and other relations through so called features is better suited to provide the flexibility necessary for combining a large variety of applications and data sources. However, GML defines the syntactical aspects of geographic data exchange by fixing low-level data types and describing what information features contain and what relations exist. Any application using a specific GML data source therefore needs to know in advance how to interpret these features and relations. This means that these relations, feature names, and resultant structure are hard-wired into the application itself. Moreover, only traditional relational queries are supported. More complex queries, incorporating transitive closures of relations, require query iterations to implement. However, such requests are commonplace in graph search algorithms such as finding a path through a navigation network.

In order to further decouple applications from this inherent structure and to support a more expressive query language, we are investigating the use of another layer of information encapsulated in a semantic network on which applications operate. The semantic network layer includes the application's view of the world and the data sources' view and is able to integrate both. As a result, visualizations can be described independently of the structure of the underlying data source.

In the remainder of this section, it will be demonstrated with a set of examples how an integrated and semantic world model enables these methods. The semantic aspect of the world model entails that such visualizations can be developed independently from the underlying data model.

4.1 Example: Gas utility company

Professionals dealing with hidden and embedded structures can be supported in their field work with the integration of administrative information with detailed models of the physical environment. For example, a worker for a gas utility company has to find and repair a leak in a gas pipe in the field.

The worker will require accurate 3D information to locate access points to the pipe where he can measure various operational parameters of the pipe in order to locate the leak. The parameters are referenced to the points and can be queried by the system on the fly from the combined spatial-semantic database. A handheld visualization device will use the stored 3D model at the same time to display the layout and placement of the hidden pipe in the ground to facilitate planning of the repair. The worker can then proceed to carry out the maintenance.

However, the displayed information is not limited exclusively to the task at hand. Any other subsurface structure in close proximity to the site is also extracted from the model and its relevance is assessed. Obstructing structures are identified by analyzing the geometric relations they have with the pipe. For example, if another pipe is above the gas pipe of interest, the visualization will include it and show any possible areas of intersection with the work path. Similarly, other structures are checked for dangers to the planned work, such as electric fields that could trigger sparks in tools. Again the visualization will highlight such structures to inform the worker of possible risks.

After the task is completed the relevant administrative information including work time, material, client information and nature of the defect are entered by the worker and automatically related to the pipe and location by the system again. The collected information can then enter the business logic workflow of the company without further overhead. The relation to the 3D model enables automatic cross-referencing of administrative data and real-world locations and artefacts. Finally such a model can be reused for information visualization in report and analysis work.

4.2 Example: Pedestrian navigation

Navigation systems for cars have become standard equipment due to accurate and affordable tracking and high-quality road maps. Similarly, one can expect that pedestrian navigation covering outdoor city use but also indoor areas will become a generic feature of future mobile systems. Various systems have demonstrated partial results in this area. However universal navigation from room to room across two buildings, streets and even city districts remains out of reach. Besides the unavailability of ubiquitous tracking, a generic model covering all levels of details involved in such a task also does not exist.

Consider a user having an appointment with a customer in an office building across the city from her current location. Her PDA queries a web service ahead of the meeting time to compute a route from her own office to the customer's. Such a route will not only include accurate driving instructions but also information on possible parking spaces in the vicinity of the destination area, the path she has to walk to get to the office building and information on how to get to the customer's office.

A display included into her sunglasses conveys navigational information as she leaves her car and walks towards the office building. The entrance she has to use is highlighted and to give some impression of the location of the customer's office, the system highlights the windows facing out of the room on the building façade. Within the building, the system directs her to an elevator and displays, or automatically selects, the floor level she has to go to. Once on the right floor, the system points her in the right direction along the corridor.

4.3 Adaptive visualization engine

Within the scenarios described above, we identified several tasks that should be delegated to an automated adaptive visualization engine. Such an engine will operate on the given world model and its meta-data to derive which elements of the model are of interest and what the appropriate style of presentation should be.

Deriving styles and transformations. The adaptive visualization engine queries the world model for the structures directly relevant to the worker's task. From the given work area it also derives other structures that intersect the planned excavation volume. Based on the attributes associated with the structures it finds, it can assign appropriate visualization styles to them.

For example, structures that lie above the designated pipe are rendered in a bright color to draw attention to them. Nevertheless, they are rendered translucent enhancing the perception of the main structure. Structures within the volume but below the pipe are rendered with darker colors as they are not as important.

Similarly, a projection of the main structure and the excavation volume to the ground model is computed to throw a “shadow” on the real ground. Such a “shadow” maps the 3D location directly to the visible surface area and delineates the required excavation area.

Displaying dangerous areas. The worker’s system should automatically identify dangerous areas or structures in the working area and notify the worker with appropriate signals. To do so it needs a model of the possible hazards related to the task at hand and a method of querying the world model to retrieve structures that fit the model.

To implement such functionality the application has a list of attributes that relate hazardous structures to the selected task. Then a query is formulated to search for structures intersecting the working area and being annotated with attributes matching the possible hazards. The query is translated into one or more WFS queries and is send to the server.

The results are converted into a scene graph suitable for rendering and interaction. Based on the returned attribute values, different levels of severity are assigned to the structures and presentation styles are set accordingly. The resulting scene graph is then added to the viewer’s graph and becomes part of the presentation.

Deriving related objects of interest. For a mobile task like navigation, objects of interest are often occluded or not fully or directly visible. In such cases it is not simple for the user to interpret the visualization correctly, even if various techniques such as transparency or cut-away views are employed. Therefore the adaptive visualization engine can substitute currently visible structures for the actual target structures and use these in the presentation.

For example, as the user is coming close to the target building both the building entrance and a window associated with the target room are highlighted. The system emphasises the entrance in order to guide the user in the right direction, but also, it draws some attention to the window to provide some overview of the planned path. The user can therefore build a mental map of where she will go.

Within the building, doorways are again used as direct navigation aids, but an outline of the intersection of the room with the adjacent walls can provide more information about the location and size of the target room. Such subtle additions to the presentation could help users to gain a better understanding of the structure of their environment.

To implement these functions, the system will first compute a path from the current location of the user to the destination. Moreover, it queries the world model for potential occluding objects in the area between the user and the destination object. If such objects exist, it will further derive visible sub-structures of these objects and try to relate them to the destination room. Such relations can be topological such as windows connecting a room volume to the outside or geometrical such as proximity of walls. If such a related object is found, it can be used in the visualization instead of the destination object.

4.4 Semantic reasoning engine

More complex automation is achieved by incorporating knowledge-based query mechanisms into the application. Our aim is to provide a software component that reasons over the available world model to identify and select objects described by more complex propositions than simple query-based assertions.

The semantic reasoning engine maintains a representation of the world and application state as a semantic network which contains information on the individual objects in the scene, the tasks of the user and the selected visualization modes. The semantic network is based on an ontological description of the properties of objects.

It combines a set of formal descriptions of application knowledge:

1. A basic ontology like OWL [HPH03] for defining the basic relationships between objects, such as membership of a class, subclasses, aggregation and attributes.
2. An ontological description of the GML-based data in order to interpret and map the GML features to the semantic network representation.
3. Application specific ontologies that further refine the basic ontology with the relations that hold in the application domain.
4. A description of the possible user interface representations and their relations to domain specific object attributes.
5. A set of mapping ontologies that translate between the GML relations and the application-specific or user interface concepts.

Finally, a knowledge base of facts about the objects is also maintained in the network. These facts are representations of the underlying GML feature data retrieved from the Web Feature Service back-end.

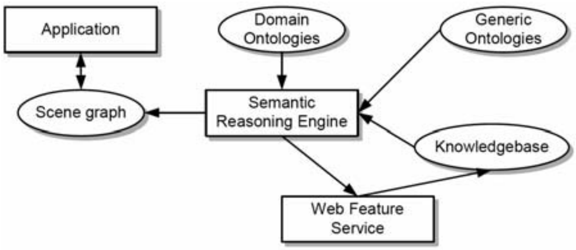


Fig. 9. The semantic reasoning engine relies on a knowledge base interpreted according to various ontologies.

An inference engine such as Racer [HM03] or FaCT [Hor98] processes queries defined by the application and returns the results with respect to the semantic network. Relations that include geometric semantics such as distance or visibility are implemented as extensions to the inference engine that operate on the geometric data itself, rather than on the representations in the semantic network. If further facts are required, the engine also formulates queries to the WFS back-end to replenish the knowledge base with more information. For example, a prerequisite to certain queries could be that all objects within a certain distance of the user are kept in the knowledge base. As the user moves, the engine will update the knowledge base as required.

4.5 Building visualizations

The next step uses the knowledge encoded in the semantic network to derive convincing visualizations. Here we use the ontology describing the properties of types of visualization. For example, we could define that objects that are occluded but important for user orientation are of type *I* and are to be rendered with a bright yellow stippled outline, while objects that are occluded by any object of type *I* are themselves of type *O* and are rendered with a black translucent color to modify the luminance of occluded objects.

The inference engine can now deduce from assertions in the knowledge base whether a certain object meets the criteria to be classified as either type *I* or *O*. Similarly, we could also formulate a query that returns all objects of a certain type. Note that the decision of whether an object is of type *I* is not solely based on geometric calculations, because it must also be important to user orientation. A formalization of this concept is in fact encoded in the domain specific application ontology.

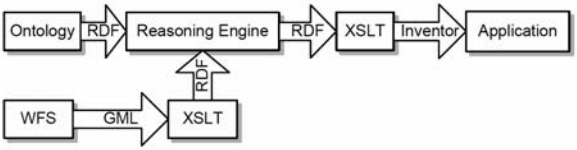


Fig. 10. XSLT is used as the general translation mechanism that interfaces the individual parts of the semantic reasoning system for AR. It can convert between geometric information (GML), semantic information (RDF) and graphical data for visualization, such as the scene graph standard Open Inventor.

The connection between the semantic reasoning engine and the user interface component is another crucial aspect of the system. The output of the engine – the results of queries to it – is transformed into scene graph representations for the rendering engine. However, due to the possible complex computations of query results, we cannot expect the reasoning engine to operate at real time in direct response to the user's interactions.

The output scene graph will rather describe the possible visualizations of all objects with respect to certain changes such as visibility or distance to the user or actions by the user such as selection. The visualization will be rendered with appropriate parameters driven by the sensor input and adapt at render time to such simple changes. Only larger movements exceeding certain thresholds or interactions with the application such as selecting a different mode or target will trigger a re-evaluation by the semantic reasoning engine.

4.6 System integration

The integration of the semantic reasoning engine into the overall system is supported by the extensive use of XML-based data exchange formats. Ontologies are expressed in RDF [MM03] and can therefore be served by web services. The knowledge base of facts in the reasoning engine is typically also expressed in RDF. Therefore XML-based transformation technologies such as XSLT can be applied to the GML query results of the Web Feature Service to extract and formulate the assertions in RDF and feed them to the reasoning engine.

The output to the application is mainly entailing the creation of scene graphs for 3D rendering. Here we have already successfully employed the same techniques to generate scene graph descriptions from XML data structures. As the output of the reasoning engine consists of facts expressed in RDF too, we can directly apply static transformations to it and create the corresponding graph structures.

4.7 Example: Tracking target objects for navigation

To demonstrate the use of the semantic reasoning engine we will describe its operation when applied to universal navigation. Here we always want to present the user with an object that is "semantically" related to the target object, even if the target is not directly visible.

Within our application specific rule base, we define a set of rules that always compute a visible place-holder for the target object. A first rule is that the best place-holder is the target object itself, if visible. Another rule states that an object belongs to the type class *IsInteresing*, if it is visible and stands in the relation *SubstituteFor* with the target object. The relation *SubstituteFor* is defined in terms of a set of logical disjunctions and conjunctions of assertions on properties of both the target and the candidate objects. Among the assertions is the topological relation *neighboursWith* meaning a spatial relation that objects are adjacent to each other. The concrete implementation in terms of testable attributes or relations depends on the world model. Therefore it is either defined in the world model itself or as a set of geometric relations *Touches* which is part of the query language for the Web Feature Service.

For example, in order to find windows that may act as substitutes for a target room, we define the relation *SubstituteFor* to include objects of a size not larger than the target, that stand in the relation *neighboursWith* to it and that are potentially visible, because the user is currently outside of the building. The semantic reasoning engine is able to map the relations to real underlying relations in the current GML data set without further interaction with the application. Therefore, the visualizations depend on high-level abstract descriptions of interesting objects rather than on the direct low-level expressions used to compute them.

The result of the reasoning step is the set of all objects belonging to the class *IsInteresting*. The result set creates a scene graph representation that includes the objects' geometry and appropriate rendering styles. Moreover, the scene graph also includes some logic to react to individual objects becoming visible or invisible as the user moves within a certain range, such that only one object is actually rendered at any point in time. When the user moves beyond this range or selects a different target, the application re-evaluates the set by updating the knowledge base accordingly and applying the reasoning again.

5 Conclusions and future work

Mobile AR has outgrown its infancy and is getting ready for early commercial deployment. We have presented a series of platforms and application prototypes developed to assess the feasibility of key technologies in mobile AR. One noteworthy aspect of our experiments is that they use probably the largest and most systematic AR model to date. From our experiences with this model and its creation process we have learned what today's 3D modeling technologies do not provide: truly flexible interpretation of the data, which makes applications independent of assumptions concerning model structure and the relations between model entities. Semantic web technology aims to overcome this problem in the domain of online information systems.

In this paper we explained why mobile AR, with its potentially large number of clients and location-based service providers, has essentially the same requirements as document-oriented semantic web applications, but in the domain of real-time 3D information. We derive a data model which allows a suitable degree of semantic reasoning for mobile AR, and describe how it can be used in practical examples. While we already have many tools in place for the implementation of such a model through our own work and the resources available in the GIS and semantic web communities, the verification of the approach with real world scenarios will be the subject of future work.

Acknowledgements

This work was sponsored by the Austrian Science Fund *FWF* (Y193). We thank Joseph Newman and Erick Mendez for proofreading and useful suggestions.

References

[Azu97] Azuma R. T.: A survey of augmented reality. Presence Teleoperators and Virtual Environments 6, 4 (August 1997), pp. 355–385.

[CMJ04] Coelho E. M., MacIntyre B., Julier S. J.: Osgar: A scene graph with uncertain transformations. In Proc. ISMAR 2004, IEEE 2004.

[FAD02] Furmanski C., Azuma R., Daily M.: Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In Proc. ISMAR 2002, IEEE and ACM, pp. 215–224.

[FMH*97] Feiner S., MacIntyre B., Höllerer T., Webster A.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban enviroment. In Proc. ISWC'97, pp. 74–81.
[FMS93] Feiner S., MacIntyre B., Seligmann D.: Knowledge-based augmented reality. Communications of the ACM 36, 7 (July 1993), pp. 53–62.
[Hor98] Horrocks I. Using an expressive description logic: FaCT or fiction? In Proc. Princ. of Knowledge Representation and Reasoning KR'98, pp. 636-647.
[HFH*01] Höllerer T., Feiner S., Hallaway D., Bell B., Lanzagorta M., Brown D., Julier S., Baillot Y., Rosenblum L.: User interface management techniques for collaborative mobile AR. Computer & Graphics 25, 25 (2001), pp. 799–810.
[HHF*04] Hallaway D., Höllerer T., Feiner S., Bridging the gaps: Hybrid tracking for adaptive mobile AR. In: Applied Artificial Intelligence, Special Edition on Artificial Intelligence in Mobile Systems, 25, 5 (July 2004).
[HM03] Haarslev V., Möller R., Racer: A core inference engine for the semantic web. In York Sure and Óscar Corcho, editors, Proc. EON2003.
[HPH03] Horrocks I., Patel-Schneider P.F., van Harmelen F. From SHIQ and RDF to OWL: The making of a web ontology language. Journal of Web Semantics 1, 1 (2003), pp. 7-26.
[JLB*00] Julier S., Lanzagorta M., Baillot Y., Rosenblum L.,Feiner S., Höllerer T.: Information filtering for mobile augmented reality. In Proc. ISAR 2000 (Munich, Germany, October 5–6 2000), IEEE and ACM, pp. 3–11.
[KG03] Kolbe T. H., Gröger G.: Towards unified 3d city models. In Proc. of the ISPRS Comm. IV Joint Workshop on Challenges in Geospatial Analysis, Integration and Visualization II (Stuttgart, Germany, 2003).
[LT04] Leykin A., Tuceryan M.: Automatic determination of text readability over textured backgrounds for AR systems. In Proc. ISMAR 2004, IEEE.
[MM03] Manola F., Miller E. Rdf Primer. W3C, http://www.w3c.org/TR/rdf-primer/, visited May 3 2003.
[NWB*04] Newman J., Wagner M., Bauer M., MacWilliams A., Pintaric T., Beyer D., Pustka D., Strasser F., Schmalstieg D., Klinker G.: Ubiquitous tracking for augmented reality. In Proc. ISMAR 2004, IEEE.
[PAT*04] Piekarski W., Avery B., Thomas B. H., and Malbezin P., Integrated Head and Hand Tracking for Indoor and Outdoor AR. In Proc. IEEE Virtual Reality Conference, (Chicago, IL, USA, March 2004), pp. 11-276.
[RCK*05] Reitmayr G., Chiu C., Kusternig A., Kusternig M., Witzmann H.: iOrb - unifying command and 3d input for mobile augmented reality. In Proc. IEEE VR Workshop on New Directions in 3D User Interfaces, IEEE, pp. 7–10.
[RS01] Reitmayr G., Schmalstieg D.: OpenTracker – an open software architecture for reconfigurable tracking based on XML. In Proc. IEEE Virtual Reality 2001 (Yokohama, Japan, March 13–17 2001), pp. 285–286.
[RS03] Reitmayr G., Schmalstieg D.: Location based applications for mobile AR. In Proc. AUIC 2003, Biddle R., Thomas B., (Eds.), vol. 25 (3) of Australian Computer Science Communications, ACS, pp. 65–73.
[RS04] Reitmayr G., Schmalstieg D.: Collaborative augmented reality for outdoor navigation and information browsing. In Geowissenschaftl. Mitteil. Vienna, vol. 66, pp. 53–62.

[SFH*02] Schmalstieg D., Fuhrmann A., Hesina G., Szalavari Z., Encarnacao L. M., Gervautz M., Purgathofer W.: The Studierstube AR project. PRESENCE – Teleoperators and Virtual Environments 11, 1 (2002), pp. 32-54.

[TDP*98] Bruce H. Thomas, V.Demczuk, Wayne Piekarski, David Hepworth, and Bernard Gunther. A wearable computer system with AR to support terrestrial navigation. In Proc. ISWC'98, IEEE and ACM, pp. 168–171.

[VZ03] de Vries M., Zlatanova S., Interoperability on the Web: The Case of 3D Geo-Data. In Proc. IADIS International Conference e-Society 2004 (Avila, Spain, July 16-19), pp. 667-674.

[WS03] Wagner D., Schmalstieg D.: First Steps Towards Handheld Augmented Reality. Proc. 7th International Conference on Wearable Computers, (White Plains, NY, USA, October 21-23, 2003), pp. 127-135.

„Anywhere Augmentation“: Towards Mobile Augmented Reality in Unprepared Environments

Tobias Höllerer, Jason Wither, Stephen DiVerdi

Four Eyes Laboratory, Department of Computer Science
University of California, Santa Barbara, California, USA
holl@cs.ucsb.edu, jwither@cs.ucsb.edu, sdiverdi@cs.ucsb.edu

Abstract

We introduce the term „Anywhere Augmentation” to refer to the idea of linking location-specific computing services with the physical world, making them readily and directly available in any situation and location. This chapter presents a novel approach to „Anywhere Augmentation“ based on efficient human input for wearable computing and augmented reality (AR). Current mobile and wearable computing technologies, as found in many industrial and governmental service applications, do not routinely integrate the services they provide with the physical world. Major limitations in the computer's general scene understanding abilities and the infeasibility of instrumenting the whole globe with a unified sensing and computing environment prevent progress in this area. Alternative approaches must be considered.

We present a mobile augmented reality system for outdoor annotation of the real world. To reduce user burden, we use openly available aerial photographs in addition to the wearable system's usual data sources (position, orientation, camera and user input). This allows the user to accurately annotate 3D features from a single position by aligning features in both their firstperson viewpoint and in the aerial view. At the same time, aerial photographs provide a rich set of features that can be automatically extracted to create best guesses of intended annotations with minimal user input.

Thus, user interaction is often as simple as casting a ray from a firstperson view, and then confirming the feature from the aerial view. We examine three types of aerial photograph features – corners, edges, and regions – that are suitable for a wide variety of useful mobile augmented reality applications. By using aerial photographs in combination with wearable augmented reality, we are able to achieve much higher accuracy 3D annotation positions from a single user location than was previously possible.

1 Introduction

A lost traveler walks down an unfamiliar street in a foreign country, looking for a specific address. His city map does not help him since all the street signs are in a foreign alphabet and he barely knows how to pronounce the street he is looking for. Fortunately, he carries an „Anywhere Augmentation“ device, which he uses as a pedestrian navigation tool to overlay position-specific directions immediately onto his field of view through his cell-phone display, which films and augments the scene in front of him. He interacts with the physical scene, pointing out cross streets, whose names, in response, appear directly on top of the physical world and are also highlighted in an optional map view of the neighborhood.

The leader of a reconnaissance team scouts out the terrain behind a hill that shields his troop from enemy view but also prevents them from surveying the landscape features and infrastructure behind it. He dons his augmentation glasses, and despite imperfect localization is quickly able to align the contours of the virtual 3D elevation model with the outline of the hill in front of him, establishing registration for overlaying the landscape and building features behind the hill directly in his field of view, in the fashion of Superman’s X-ray vision. He makes the decision on how to approach that terrain in a much more informed fashion.

A schoolchild is on a field trip to learn about botany. Distinguishing different orders and families of trees has never been his strength, but help is at his fingertips. His small electronic companion allows him to see labels of already classified trees directly overlaid on his view of them and allows him to add tentative new classifications and take pictures for later verification by the teacher or field guide.

What these three scenarios have in common is the idea of having computational resources available anytime and anywhere, and moreover, being able to form a link between the location-specific information and the physical world by direct vision overlays. We introduce the term

„Anywhere Augmentation“ for this concept, emphasizing the necessity for such a system to work in arbitrary environments in order to become adopted by users and make life easier for them. Currently, such technologies exist only for very limited example applications in research laboratories.

Mobile and wearable computing technologies have found their way into mainstream industrial and governmental service applications over the past decade. They are now commonplace in the shipping and hospitality industries, as well as in mobile law enforcement, to highlight but a few successful examples. However, current mobile computing solutions outside of research laboratories do not sense and adapt to the user’s environment and they do not link the services they provide with the physical world. And because of major limitations in the computer’s general sensing and scene understanding abilities and the infeasibility of instrumenting the whole globe with a unified sensing and computing environment, this is not likely to change soon, unless we find alternative approaches – which is the starting point for our work.

2 From Mobile Augmented Reality to „Anywhere Augmentation”

In spite of the great potential of mobile AR for many application areas, progress in the field has so far almost exclusively been demonstrated in a number of research prototypes. Actual commercial deployment is limited; early commercial technology and service providers are struggling to find customers and create markets. Despite better solutions to the technical challenges of wearable computing problem areas remain, such as the need for miniaturization of input/output technology and power supply, and for improved thermal dissipation, especially in small high-performance systems. Also, ruggedness is a key requirement. Outdoor AR is particularly challenging; in contrast to controlled environments indoors, one has little influence over outdoor conditions. Lighting can range from direct sunlight in a reflective environment to absolute darkness during the night. Outdoor systems should withstand all possible weather conditions, including wind, rain, frost, and heat.

However, it is not chiefly because of these issues that „Anywhere Augmentation“ has not yet emerged as a widely adopted application technology. After all, many people operate their cell phones or their MP3 players comfortably, even under adverse conditions. It is already possible to manufacture hardware that can function in all the environments that we have in mind for „Anywhere Augmentation“. First, however, it has to be demonstrated that these devices can be usefully employed, and the problem is, at least to a certain extent, one of user interface design. With standard graphical user interfaces straightforwardly adapted from desktop and laptop computers, we have not seen sufficient user enthusiasm to warrant launching a whole new wave of wearable and situated computing computing products.

AR, which makes the physical world a part of the user interface experience, has the potential to play a significant role in changing this.

One of the main problems with current approaches to augmented reality is that in order to obtain reliable and accurate registration between the physical world and the augmentations one either needs a model of the environment, or the environment needs to be instrumented, at least passively with registration markers. Both of these preconditions severely constrain the applicability of AR. Instrumentation of environments on a global scale is exceedingly unlikely to take place, and detailed 3D city and landscape models are very cumbersome to create [1]. In addition, even if detailed 3D models of target environments existed on a broad scale, keeping them up-to-date would be a major challenge and they would still not take into account dynamic changes. Instead of relying on the existence of data that is not likely to become available in the near future, we propose to utilize several sources of GIS data for which there are already data repositories with nationwide coverage (e.g. aerial photography, elevation, land use, street maps, NGA names database). The general concept of „Anywhere Augmentation“ does not depend on the existence of any particular subset of these data sources, but instead we want to consider any of these sources when available, and their existence will improve the user experience by providing more information and stronger constraints for user interaction.

Annotation of outdoor scenes is an important part of mobile augmented reality research (cf. Fig. 1). Generally, the situated content displayed by a wearable system is carefully constructed offline using many different technologies, including modelling programs, GIS data, and aerial photographs. In this work, our focus is on annotating an outdoor scene from within the wearable system, providing an appropriate interface to allow accurate markup in a mobile context. To reduce the amount of manual work that must be done by the user, we have modified our system to use aerial photographs of the region in conjunction with the wearable’s acquired data. This allows the user to accurately place 3D annotations from a single position by providing a means of accurately gauging depth.

Fig. 1. A wearable system for outdoor annotation. Left to right: (a) A user wearing the system to annotate an outdoor scene. (b) The user's first-person view, showing the scene from the ground and the visible annotations. (c) The user's overhead view, showing an aerial photograph of the local region with the user's position and placed annotations overlaid.

With orientation tracking, from a static position a user can easily cast a ray to select a visible feature in the scene, but setting the depth of that feature is more difficult. Previous work in this area requires the user annotate the same feature from multiple viewpoints to triangulate a position [22], or estimate depth from a static viewpoint using artificial depth cues [31]. However, commonly available aerial photographs [14, 32] can be used to allow accurate 3D position input from a single location. After a user has cast a ray, our system presents the user with an aerial view of the scene and the cast ray and allows the user to adjust the ray and set a distance. The result is a significant improvement in the accuracy of 3D positions over previous AR distance estimation work [31], as well as the ability to annotate features that may not be directly visible from the user's location, such as the opposite side of a building. Automatic feature extraction from the aerial photographs allows the system to intelligently recommend salient features along the cast ray, so the user needs only to choose from the detected features and possibly refine the result.

We examine three different types of features a user may want to place in the outdoor scene, based on how they appear in the aerial photograph. Corners can correspond to the vertices of building silhouettes and are useful for modelling geometry [4]. Vertical walls appear as edges that can be used to properly orient and position world-aligned billboard annotations [15]. Uniform regions in aerial photographs can denote buildings, fields, etc. and can be annotated with a label and a bounding box for wearable navigation purposes [12]. Our manual interface and automatic feature extraction techniques are thus geared towards finding these types of features in our aerial photographs. We use these annotations as a representative set of possible information a user may want to input, but our system is not geared towards any particular application and only minor modifications are needed to tailor the approach to other task scenarios.

A key focus of this work is the aggregation of available data sources, in this case the wearable's data streams and the aerial photographs, to reduce the burden on the user for traditional AR tasks. This is the first step towards our goal of anywhere augmentation, where the usual AR initial costs of manual modelling, calibration and registration are alleviated to make augmented reality readily available in any unprepared environment. Our contribution is to significantly reduce the work necessary to create physically-situated annotations in an unprepared, large-scale outdoor scene. The development and use of real-time, local, automatic feature extraction techniques for aerial photographs is a secondary contribution of this work.

3 Previous Work

The previous work for this project can be split between our two contributions. The first section compares our approach with other wearable systems dealing with outdoor annotation. The second section discusses feature extraction from aerial imagery.

3.1 Wearable Systems

Rekimoto et al. [25] introduced the idea of Augment-able Reality with a system that allows users to annotate the environment with contextual information at specific locations that had been prepared ahead of time with active or passive markers. They envisioned extending the system to allow annotations for any position of known GPS coordinates. Our system expands on this concept by allowing annotations of unprepared environments at arbitrary locations without known GPS coordinates.

More recent work has been done in using wearable systems to acquire accurate positions of arbitrary locations, towards the goal of modelling outdoor scenes from within a wearable system. Baillot et al.'s [4] wearable modeller is based on first creating 3D vertices by casting intersecting rays from multiple viewpoints, and then creating primitives fit to those vertices for the final model. Our annotations are a more easily acquired and more accurate version of their construction points, and could be used for the same sort of modelling application as they describe. The paper also shows an example using an indoor scene's architectural floor plan as a guide for creating vertices. This is the inspiration for our use of aerial photographs, but we extend the concept in many new directions – we use commonly available aerial photographs that are automatically registered to the user's location, and we use them not only for creating points, but for many other types of annotations as well, and we can even use them to create accurate annotations for features that are not visible from the current location, such as placing a correctly oriented billboard label on the opposite side of a building.

Piekarski and Thomas' outdoor wearable modelling system [21, 22] implements a wide variety of techniques based around the concept of working planes that are created by sighting down the wall or feature to be modelled from one viewpoint, and then moving to another location to create points on that plane at the correct depth. Our solution replaces the need for working planes with the overhead view – working planes the user would normally have to construct from a particular location are already available as edge features in the aerial photograph, removing the need for the user to move around large buildings to distant locations for accurate modelling.

To avoid requiring multiple viewpoints for determining 3D positions, Reitmayr and Schmalstieg's system [24] has a complete model of the environment (obtained offline) that users can annotate by casting rays that intersect with the model's surfaces. The goal of „Anywhere Augmentation" aims to remove the initial start up costs associated with acquiring a detailed scene model offline – in our system, we remove the need for such a model by using aerial photographs to provide the same features for annotation.

Maps and aerial photographs have also been used in many mobile systems for passive localization purposes. The Cyberguide system [2] uses a hand-held device to display a rough estimate of the user's position and orientation on an abstract map with point-of-interest annotations. ARVino [17] uses aerial photographs in a virtual reality view of GIS data, to aid the user in mentally mapping abstract information onto the physical environment the data annotates. We extend the functionality of both Cyberguide and ARVino by using aerial photographs for passive localization of the user, as well as active annotation of the environment by the user and even for automatic extraction of new features.

Reitmayr et al. extract features from paper maps to display registered annotations with an overhead projector [23]. The type of features they use differ from ours in that they are geared towards robust identification and registration for displaying overlays, as opposed to our system which extracts user-specified semantic features as targets for annotation.

Finally, in place of a map, Bell et al. [7] and Güven and Feiner. [15] use a virtual model of the environment for localization by displaying a world in miniature view. This gives the user a more informed understanding of their surroundings, and could even be used to help users create situated annotations. However, detailed offline model construction is not in the spirit of our goal of „Anywhere Augmentation“.

3.2 Feature Extraction

There is an extensive body of research in the realm of feature extraction from aerial photographs for scene understanding. Mayer [19] presented a thorough survey of automatic object extraction from aerial photographs, with an emphasis on buildings. This survey compares a wide range of projects in terms of their final output – while many of them create detailed building geometries, there are as many projects that focus on low complexity geometry and feature sets. A common theme of the more complicated papers surveyed is the use of sophisticated input data, such as two more images from multiple viewpoints [13], laser range data [3], or digital surface models [30]. Because of our emphasis on „Anywhere Augmentation“, we restrict our input to commonly available datasets, such as the single image, visible spectrum, near-orthographic aerial photographs from Google Maps [14] or Yahoo Maps [32].

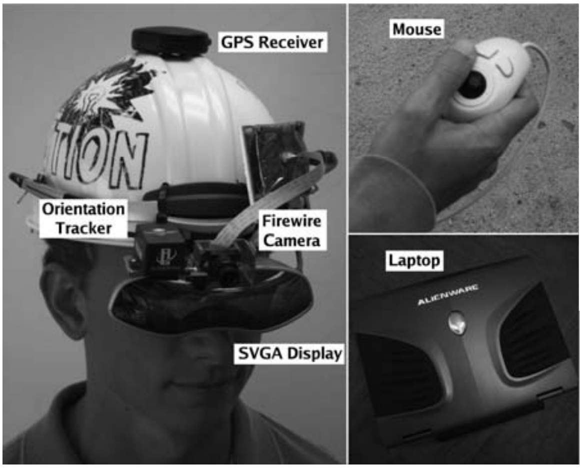


Fig. 2. Our wearable system hardware. An Alienware Area-51 m5500 laptop (worn in a backpack), an SVGA Sony Glasstron PLM-S700E display, a Point Grey Firefly firewire camera, An InterSense InertiaCube2 orientation tracker, a Garmin GPS 18 receiver, and an ErgoTouch RocketMouse.

The common trade off in automatic feature extraction is between lower complexity of input data and a higher computation cost. Many algorithms [28, 27, 18] rely on global analysis of region imagery and iterative minimization approaches that require significant compute time. Our goal of „Anywhere Augmentation“ discourages lengthy preprocessing on our datasets, which means that our feature extraction must be done on the fly, making costly algorithms impractical.

One observation that has proven useful in feature detection in a number of cases is that aerial photograph features are often correlated across different scales. Baumgartner et al. [6] use this property to search for different features of roads at different scales – lines at coarse resolutions, and uniform patches at fine resolutions – to generate an overall better model of road geometry. A slightly different approach was taken by Reitmayr et al. [23], by searching for the same features at multiple scales, taking advantage of the multiscale self-similarity of map data. Our corner extraction technique also utilizes multiscale detection based on the assumption that building silhouettes will result in salient corners at multiple scales, while noise and texture corners will not.

The performance of automatic building extraction algorithms, in terms of both speed and robustness, can be greatly improved by small amounts of user input. These semiautomatic approaches can require user input at the outset to guide detection, or after detection has occurred to correct errors. Vosselman and Knecht [29] and Kim et al. [16] both take the former approach towards the application of labelling roads for mapping. The user annotates a small section of road with the direction the road continues, which the algorithms are able to use to determine characteristics of the road and follow its path throughout the rest of the aerial photograph. Nevatia and Huertas [20] let the user correct errors in the results of their automatic building detection, and use the corrections to guide later detection steps. Our semiautomatic interface requires user input to determine the initial detection parameters, and then allows the user to manually correct errors in the detected results if necessary.

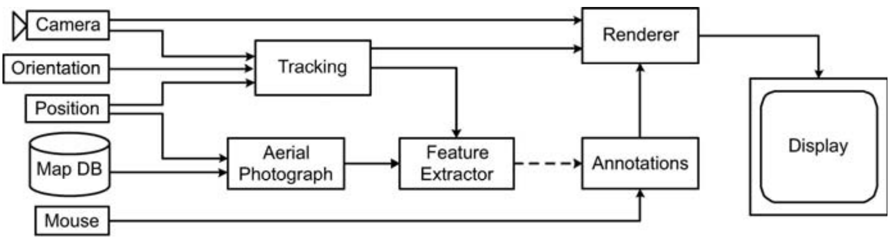


Fig. 3. An overview of our system structure. The camera, position, and orientation data are combined for a general tracking result. Position data is also used to query a database (Google Maps [14]) for the local aerial photograph. The aerial photograph and tracking data are used for feature extraction, which is optionally sent along with mouse input to the annotation model. Annotations, tracking, and the scene image are all used to render a final visual to the display.

4 System

Our wearable system can be seen in Fig. 2. At the core is an Alienware Area-51 m5500 laptop, which is worn on the user's back. The display is an SVGA Sony Glasstron PLM-S700E hanging from the front of the helmet, used in video see-through mode. Mounted directly above the display are a Point Grey Firefly firewire camera and an InterSense InertiaCube2 orientation tracker, and on top of the helmet is a Garmin GPS 18 position tracker. User input is through a hand-held ErgoTouch RocketMouse. All of these devices are relatively inexpensive, off-the-shelf components.

Many services now exist on the internet to provide access to aerial photographs and high-resolution satellite imagery of the world, including Yahoo Maps (at 1.0m resolution for the entire United States) [32] and Google Maps (with variable resolution) [14]. In our current system, we acquire 0.5m resolution aerial photographs offline from Google Maps and stitch them together into a single large view of the University of California, Santa Barbara campus. However, it is possible with a wireless internet connection to download map data on the fly based on the wearable's reported GPS coordinates. This automatic map acquisition would allow the system to work in new environments covered by map services without the initial setup cost.

See Fig. 3 for an overview diagram that shows all the components and connections of the complete system.

4.1 Calibration

While our position and orientation sensors report absolute coordinates in a global reference frame, both have too much error to be used without further calibration. We found that our GPS tracker can be off by as much as 10 meters, but that it also drifts slowly. To compensate for the initial offset users are asked to specify their exact location on the aerial photograph, which greatly reduces the position error for a single run of the application.

Our orientation sensor is designed to provide orientation information from true north, but has two major sources of error. First, the difference between true north and magnetic north is a systematic error, which we overcome by adding a second step to the calibration process. In this step the user centers the view at a distinct feature and then clicks a button to bring up the aerial photograph. The user's position and orientation are overlaid on the photograph, and the user is able to modify the displayed orientation to directly coincide with the user's chosen feature. This calibration procedure is similar to the single point calibration technique formalized by Baillot et al. [5], except that in our system the calibration location does not have to be predefined because the user chooses the necessary points on the aerial photograph during the process. However, because the user input is limited to adjusting the orientation in the overhead view, only error in yaw is accounted for. Roll and pitch can be roughly corrected by assuming the user's head is level and oriented vertically during the calibration.

The second source of error in the orientation tracker comes from nearby ferromagnetic materials. We found that this can distort our tracking results by as much as twenty degrees. To compensate for this, we integrated a modified version of our hybrid tracking system [31]. This system, which is based in part on previous work by Satoh et al. [26] and You et al. [33], corrects the orientation error by using gradient descent to find the best match between a set of points whose pose is updated by the inertial tracker and a matching set of points whose screen coordinates are tracked using a Lucas-Kanade optical flow feature tracker from OpenCV [10]. Individual image points are found via a corner finding algorithm from OpenCV, and the matching inertial points are computed by unprojecting the screen coordinates to a set distance from the user.

5 Annotation Interface

The intuition behind our use of aerial photographs to assist annotation is that they can fill the role of the second point of view necessary for triangulating 3D positions. Rather than having to walk to a different location and view the point again to find its depth, the user can instead find the point on an aerial photograph to provide the necessary information to calculate the distance to that point. This also provides a unified way of making many different types of annotations by marking up an aerial photograph – for example, specifying a corner, an edge or a region all correspond to well-understood 2D drawing tasks on an aerial photograph. Additionally, tasks such as specifying the perimeter of an entire building or labeling the rear wall of a building can be possible using aerial photographs, whereas they would otherwise require moving around to the building's opposite side. Our aerial photographs are always shown in a north-up orientation which has shown to be faster than a forward up configuration for search tasks like ours where the target is not shown on the aerial photograph [11].

While aerial photographs provide many opportunities for annotations by themselves, they are especially useful in combination with a wearable system. With only an aerial photograph, it is possible to annotate many types of features, but only in 2D – modelling the accurate height of a building would be very difficult. Our wearable system provides an advantage by allowing the specification of a 3D position from the combination of the aerial photograph with the first person viewpoint. The usefulness of aerial photographs is also greatly increased when the user can be situated in the environment they are annotating. For example, it may be difficult to distinguish features in an aerial photograph alone, but when a user can stand in the scene and look at the buildings from a ground-level viewpoint, these ambiguities can be more easily resolved. Having both the aerial photograph and the first person view is analogous to having a perspective view and a top-down view in a CAD modeller – while many things can be done in either view independently, having both views is often faster and more powerful.

Our annotation system utilizes these two views of the scene to make specifying annotations very easy. First, the user casts a ray in the direction of the feature to annotate by centering it in the field of view in first person mode. Once a ray has been cast, the view switches to an aerial photograph mode, with the user’s position and the ray overlaid. Then only a few simple interactions are necessary for the user to create any type of annotation at the correct location in the overhead view.

The interaction techniques we chose to use for annotating are focused on the ray that is cast from the first person view. To place an annotation, the user casts a ray in the appropriate direction, and then sets the distance of the annotation along that ray. We chose to break this interaction into two one-dimensional tasks to reduce user burden and increase the resulting precision in the less-precise wearable environment [8], and to keep the ray cast by the user central to the annotation interaction. Instead of using mouse input, it would also be possible to interact with the aerial photograph using a tablet PC. However, it has been shown that while stylus interactions are faster in a wearable environment, they are less precise as well [9].

The three types of annotations we examine are corners, edges and regions. See Fig. 4 for examples of each of these features in the aerial photograph view. The specifics of the interface for each type of annotation, as well as example applications for each type of annotation are described in the three sections below.

5.1 Corners

The most general type of feature, corners can be used for many different applications. They are not limited to corners of buildings but can represent any feature that has a distinct, visible location in the aerial photograph. This could include objects like light poles along a street, doors of buildings (if paths lead to them), trees, or even features on the ground like street or sidewalk corners.

The most straightforward application for corner annotations is modelling. These corners could be used like Baillot et al.'s construction points [4], or as a sparse point cloud in any other modelling application. Corners could also simply be used as 3D points to bind contextual information to.

Fig. 4. The aerial photograph view from within the system. In the lower left corner, an insert of the video feed from the head-worn camera can be seen. The user's position and orientation are represented on the photograph with a small cone avatar. A small set of features have already been annotated – two corners (the green points), one edge (the green line), and one region (the transparent green rectangle).

Placing a corner in 3-space is a three step process in our system. First the user finds the feature they want to annotate and centers that feature in their field of view. A mouse click changes to the aerial photograph view, with the user's position and the cast ray drawn on top. If tracking error has caused the ray not to intersect the selected feature, the user can change its direction by rolling the trackball in the direction the user wants the ray to move. When satisfied, another mouse click creates an annotation on the ray at the user's position. Rolling the trackball away from the user moves the annotation further away along the ray – once the annotation is at the same distance as the feature, a final mouse click completes the annotation. The view returns to first person mode and the user can see their corner annotation as a small cube (see Fig. 5a). An important note here is that the annotation will appear at the correct height in the first person view, because the user originally cast a 3D ray in the first step.

5.2 Edges

Edges in aerial photographs are useful for many different kinds of annotations. Multiple edges could be used to model anything with sharp image boundaries, such as building perimeters, fields, pools, sidewalks and roads. Another use for annotating edges is placing world registered billboard labels. For instance Güven and Feiner [15] use world-stabilized images in their authoring environment to localize the information they are presenting, by displaying the annotation as a billboard on an existing structures.

Fig. 5. Example annotations as seen by the user in first person view mode. Left to right: (a) Corner annotations on the corners of two buildings are rendered as cubes. (b) An edge is annotated with a texture mapped onto the plane of the wall it denotes. (c) A region annotation is rendered as a wireframe bounding box. These renderings are not geared towards a particular application; rather, they are for illustrative purposes. Applications using these annotations would have visual representations tailored to their needs.

Creating an edge annotation in our system follows the same basic procedure as creating a corner annotation. The user centers the edge to be annotated in first person view, adjusts the cast ray and sets the distance along the ray. Instead of a point, the edge annotation is drawn as a line segment perpendicular to the cast ray. Once the annotation is positioned correctly, a final step is needed to adjust its orientation to align with the feature being annotated. This is done in the same way the cast ray is adjusted, by moving the trackball in the desired direction of rotation. After the annotation is fully specified, the display returns to the first person view where the user can see the new annotation (see Fig. 5b).

5.3 Regions

Regions are the third type of feature we have chosen to annotate. An example of why region annotations are useful is given by Feiner et al. [12], who use screen oriented, world stabilized annotations to label buildings. These annotations can be particularly useful if tracking is not robust enough to support more tightly registered annotations such as edge or corner annotations. We also give our region annotations a width and depth (the x- and y-axis on the aerial photograph, respectively), so regions can very quickly be used as a rough axis-aligned model. This is obviously most useful when the buildings are also rectangular and axis-aligned. Then, the user can get a simple but complete model by casting a ray towards the roof of a building to give the annotation the appropriate height as well as width and depth. Generally, any large aerial photograph feature could be usefully annotated by regions, such as fields and parking lots, or even more visually complex semantic regions like a park full of trees or a group of buildings.

Specifying a region annotation follows the same basic steps as the corner annotation. This time, the user casts a ray through the center of the area to annotate and sets the distance along the ray so the final position is at the center. To finish the region, the user then drags out a corner of the bounding box to fit the area on the aerial photograph, and that action is mirrored for the other three corners of the bounding box. The result is the region annotation bounds the area in the aerial photograph, and its height is set to the height of the original ray at the distance to the annotation's center. Afterwards, the display is returned to the first person view, where the region annotation is drawn as a wireframe box (see Fig. 5c).

6 Feature Extraction

In addition to providing a useful viewpoint for users to manually annotate an outdoor scene, aerial photographs also provide a great deal of information that can be automatically segmented with appropriate image processing. We leverage this information by attempting to automatically detect the feature the user is annotating. If the feature is detected correctly, the user only needs to confirm it in the overhead view – otherwise, the same selection interface can be used to correct any errors in the detected feature. Thus, the semiautomatic approach does not significantly add complexity to the interface, but does frequently reduce the amount of input necessary, significantly reducing the burden on the user.

Fig. 6. The region of the aerial photograph searched for features. The user's view is represented by a small cone avatar. The white rectangle is the local portion of the aerial photograph the filters are applied to, and the dotted red lines show the region searched for for valid features. The angle swept by the dotted red lines is equal to twice the expected orientation error.

The main limitation of our automatic feature detection is that it must be fast, as it is executed each time the user casts a ray. To reduce the amount of work required each step, the detection algorithm is only run on a small search region that contains the ray cast from the user's position out to a maximum distance, rather than the entire aerial photograph (see Fig. 6). This reduces the amount of data to be processed at any given time by an order of magnitude. We also restrict the feature detectors to use only fast, local filters, instead of global or pairwise-pixel operators. Given these constraints, performance of the filters has not been a problem for the user experience.

Small errors in the tracking and the imperfectness the feature detection necessitate a certain amount of flexibility in the set of detected features returned. To address orientation tracker error, an angular epsilon term is used to define a search cone around the cast ray in which valid features may be found (see Fig. 6). For the case that the best detected feature is not the intended feature, the best n features are returned.

The annotation interface must be slightly modified to support automatic feature detection. After the user centers a feature in the field of view and casts a ray, in the aerial view of the scene, multiple detected features along the ray are presented. The user can then easily scroll through the options in the order of distance from the user by rolling the trackball up or down to select one. If the selected annotation is accurate enough, the user confirms it by clicking the middle mouse button and the view returns to first person mode. Otherwise, the user clicks the right mouse button and is presented with the same interface to adjust the annotation as for manual positioning.

6.1 Corners

Because of the relatively large size of buildings with respect to other features in an outdoor environment, building corners have the convenient property that they are persistent at a large number of different scales of the same image, whereas noise corners that come from texture, image noise, or small objects, disappear at coarser scales. On the other hand, at coarser scales, some pixels may combine to create new false corners, and small, distinct objects with valid corners may be lost entirely. Therefore, a multiscale approach to corner detection will provide more robust results for all sizes of corners a user may want to annotate. Our basic approach is to generate a corner image of the local region the user is annotating, where each pixel represents the likelihood that that pixel is a corner, by using OpenCV's Harris corner detector the region at multiple scales on multiple scales and summing the results (see Fig. 7a). Local maxima of the smooth corner function are extracted by a sliding 5x5 window. Then, the region along the user's cast ray is searched for the maximum weight pixels (from the set of local maxima). The weighting function is

$$w = w_s * w_a * w_d \tag{1}$$

where ws is the strength of the corner sampled from the corner image, wa is an angular term and wd is a distance term. wa is computed by finding the angular offset to a pixel from the cast ray and interpolating the weight from one to zero as the offset goes from zero to the angular epsilon. wd is determined by the distance to the pixel – it is set to one past a minimum distance threshold, and interpolated between zero and one within the threshold. An example set of detected corner features can be seen in Fig. 7b.

6.2 Edges

Finding edges in the search region is a simple matter of using OpenCV's Canny edge detection operator. However, obtaining the most salient edges from that image is more complicated. Our approach is to find the set of connected contours in the edge image and use OpenCV's polygon approximation algorithm to simplify the contours to within a certain error threshold. After the simplified contours are generated, all segments below a minimum length threshold are removed. The minimum length is linearly interpolated between a threshold for dominant-direction edges (0 or 90 degrees), and a different threshold for diagonal edges (45 degrees off the dominant directions). This is based on the observation that buildings tend to have 90 degree corners and tend to be oriented in the same direction as one another – for our example dataset, most of the buildings are axis-aligned – whereas noise edges tend to be randomly oriented. If there is no correlation among building orientations, the two thresholds can be set to the same value, removing the bias.

Once the final set of edges is determined (see Fig. 7c), the weight of each edge is calculated and the maximum weight edges are returned. The weighting function is

$$w = w_i * w_o * w_d * w_g \quad (2)$$

where wi is an intersection term, set to zero if the edge does not penetrate the cast ray's error region, interpolated up to one if the edge intersects the actual cast ray. wo is an orientation term, interpolated between one if the edge is perpendicular to the cast ray, and zero if the edge is parallel. wd is the same distance term as used for the corner detection – if the edge is within a minimum distance threshold, it is weighted lower. The last term, wg is a strength of the edge, determined by sampling the magnitude of the image gradient at the edge's midpoint. The gradient is computed using OpenCV's Sobel filter along the x and y dimensions. An example of the detected edges in a region can be seen in Fig. 7d.

6.3 Regions

As region features are areas like building rooftops, fields, parking lots, etc., they often appear in aerial photographs as areas of uniform pixel values plus some local texture, surrounded by a boundary edge. To find regions then, the first step is to attempt to reduce the influence of local texture by using OpenCV's closure morphology operator. The Canny edge detector is then applied to the texture-suppressed image to find region borders. Ensuring connectedness of the boundaries is necessary, so the closure operator is applied to the edge image, and then thinning is done to restore single-pixel wide edges. The result is a binary image that segments the aerial photograph into regions of similar color. The components are extracted using a flood fill algorithm, and basic component metadata is computed including average color and bounding box. This tends to produce many small components, with buildings and fields split up across multiple regions. To combine these small components we first cull some components based on HSV representation – since we focus on buildings, grass and tree areas are distractions, so any components with a basic green appearance are thrown out (this application-specific simplification could easily be replaced by something more sophisticated such as a clustering approach to vegetation segmentation). Then components are combined based on the percentage overlap of their bounding boxes and their perceptual color similarity as calculated by the euclidean distance in CIE L*a*b* color space, and components that are below a minimum area threshold are discarded. The final list of components (see Fig. 7e) is then weighted and the maximum weight components are returned.

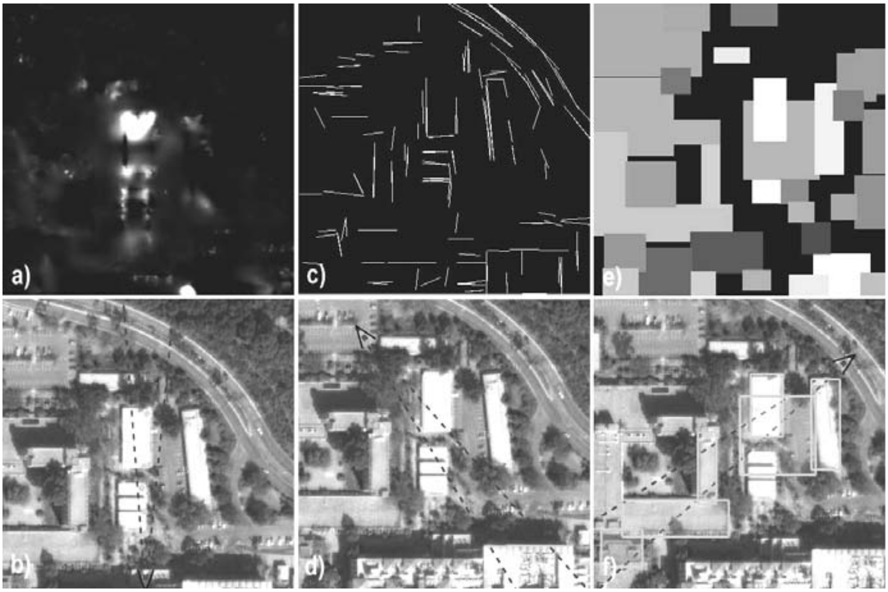


Fig. 7. Outputs of each of the automatic feature detectors. Top to bottom, left to right, : (a) The Harris corner transform is applied at multiple scales and summed together. (b) Corner features are selected from the local maxima of the continuous function. (c) Coherent line segments extracted from the output of Canny edge detection. (d) Edge features are selected from these line segments, weighted by their gradient magnitude. (e) After segmenting uniform color regions, components are merged and represented by their bounding boxes. (f) Region features are selected from components by size, intersection and distance from the center.

The weight function is

$$w = w_i * w_a * w_p * w_d \tag{3}$$

where wi is a binary intersection term – if the ray is cast from within the region, or if the ray does not intersect the region, it is set to zero, otherwise one. wa is an area weight term, interpolated from zero to one between a minimum and maximum area value. wp is the perpendicular term, calculated as the perpendicular distance between the center of the region and the cast ray, as a percentage of the region's diagonal length. wd is the same distance weight term as the corner and edge detectors, penalizing regions that are too close to the user. A typical set of detected region features can be seen in Fig. 7f.

7 Discussion

Informal testing shows that the use of aerial photographs allows users to annotate scene features in 3D from a static viewpoint with much greater precision than was previously possible [31]. The longitude and latitude accuracy of an annotation position is limited only by the accuracy of the map and the ability of the user to manipulate the position accurately. Google Maps provides data at 0.5m per pixel resolution for Santa Barbara [14], and user input is generally accurate within a few pixels, so our final annotation precision is §1.5m. While additional interface modifications may make pixel-accurate user input possible, it is unlikely precision would increase as expected since photograph data often has noise and blurring that make sharp features like corners and edges appear to occupy multiple pixels. For some feature types, automatic, local energy minimization could potentially provide subpixel accuracy. Since height information is computed from the ray cast by the user, its accuracy is dependent on the quality of the orientation tracking.

The performance of the automatic feature extraction was informally tested in an offline environment. For each type of annotation, 7 user positions were selected from a large area aerial photograph, and for each location, multiple visible features were targeted to annotate (57 corners, 34 edges and 31 regions). The detected features were inspected, and if any were close enough to the intended feature that manual correction would not be necessary, it was recorded as a success. The results of these tests were that corner detection was successful approximately 65% of the time, as was edge detection, while region detection had a success rate of approximately 40%. We want to make very clear that even a 40% success rate leads to a substantial speedup of user interaction in the general case since there is no considerable penalty to pay for “failed” feature preselection. In the worst case, the user simply resorts to completely manual selection.

Once annotations are placed, the accuracy of their appearance in the first-person augmented reality view is determined by our system’s tracking accuracy. Even with our hybrid vision and gyroscopic orientation tracking, there is still error up to five degrees from the automatic acquisition and reintroduction of new vision features. Standard PC GPS units make cheap, wide-area position tracking possible, but at low accuracy. Our system regularly experiences drift of up to a few meters over short periods of time even in clear conditions. Differential GPS or a hybrid GPS and vision or inertial position tracker would improve this result. These small position and orientation errors result in apparent mismatch between annotations and image features in first-person view, even with good annotation position accuracy.

Aerial photographs bring with them a number of limitations. First of all, while we use nearly top-down orthographic images, there is still a slight off-axis view angle that causes the roofs of tall buildings to shift a small, non-uniform distance (up to five pixels) from their ground perimeter. Currently, we do not account for this effect. Actual top-down orthographic aerial photographs or more sophisticated image processing would alleviate this problem.

The resolution of aerial photographs limits what sort of features they are useful for annotating. Small objects such as lamp posts, flag poles, or picnic tables may occupy too few pixels to extract multiple corners, or may even not appear recognizably at all. Higher resolution aerial imagery is steadily becoming more available and will help address this problem. However, aerial photographs are also captured infrequently, sometimes with many years elapsing between updating regions. This means that non-architectural objects such as picnic tables, cars, bike racks, temporary installations will not be represented or will be represented inaccurately. More troublesome can be new buildings that do not appear at all in a photograph.

8 Conclusion

We present a novel mobile augmented reality system for the annotation of features in large, outdoor scenes. Our primary contribution is the integration of a new data source, aerial photographs, to significantly reduce user burden while increasing annotation accuracy. Our secondary contribution is the use of real-time, heuristic automatic feature extraction on aerial photographs to further reduce user burden. The end result is a significantly improved interface and user experience for the traditional augmented reality task of outdoor annotation.

There are many avenues for improvement in future work. Foremost is the opportunity to include additional data sources, such as elevation data, road maps, and other GIS data. Fusion of many different, commonly available sources will improve the robustness and general applicability of our technique. Examining a larger array of annotation types would allow for a more complete model of the scene with more general usefulness in a wider variety of applications. A greater level of sophistication in the automatic feature extraction would further reduce the user burden and could improve the overall time required to model a scene. Finally, better tracking technologies to stabilize GPS drift and reduce orientation inaccuracy would enhance the user experience in first-person mode after annotations have been placed. We are currently developing a cheap tracking modality with no significant setup requirements based on a small camera pointing towards the ground so as to analyze optical flow in the way an optical mouse does on a smaller scale. By itself, this yields high frequency, high resolution relative position information similar to an inertial navigation system, but with significantly less drift. When coupled with a wide area tracking modality via a complementary kalman filter, the hybrid tracker becomes a powerful base for indoor and outdoor mobile mixed reality work.

Most important for us among all these future work opportunities is to remain true to the goal of „Anywhere Augmentation“. The work must continue to emphasize low startup cost and quick, easy integration to new scenes. This way, the traditional barriers to high quality augmented reality can be overcome, significantly increasing the general appeal of augmented reality solutions.

Acknowledgements

Special thanks to Ingrid Skei and John Roberts for their work on the initial prototype. This research was in part funded by a grant from NSF IGERT in Interactive Digital Multimedia #DGE-0221713, and an equipment donation from Microsoft.

References

1. Mahdi Abdelguerfi. 3D Synthetic Environment Reconstruction. Kluwer Academic Publishers, New York, NY, June 2001.
2. G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: A mobile context-aware tour guide. In Proceedings of Wireless Networks, pages 421–433, 1997.
3. P. Axelsson. Integrated sensors for improved 3D interpretation. International Archives of Photogrammetry and Remote Sensing, 32(4):27–34, 1998.
4. Y. Baillot, D. Brown, and S. Julier. Authoring of physical models using mobile computers. In Proceedings of the International Symposium on Wearable Computers, pages 39–46, 2001.
5. Y. Baillot, S. Julier, D. Brown, and M. Livingston. A tracker alignment framework for augmented reality. In Proceedings of the the International Symposium on Mixed and Augmented Reality, pages 142–150, 2003.
6. A. Baumgartner, C. Steger, H. Mayer, and W. Eckstein. Multi-resolution, semantic objects, and context for road extraction. In Proc. of Semantic Modelling for the Acquisition of Topographic Information from Images and Maps, pages 140–156, 1997.
7. B. Bell, T. Höllerer, and S. Feiner. An annotated situation-awareness aid for augmented reality. In Proceedings of User Interface Software and Technology, pages 213–216, 2002.
8. D. Bowman. Interaction Techniques for Common Tasks in Immersive Virtual Environments. PhD thesis, Georgia Institute of Technology, Atlanta, 1999.
9. A. Chamberlain and R. Kalawsky. A comparative investigation into two pointing systems for use with wearable computers while mobile. In Proc. of the International Symposium on Wearable Computers, pages 110–117, 2004.
10. Intel Corporation. Open Source Computer Vision Library Reference Manual. Intel Corporation, 2000.
11. R. Darken and H. Cevik. Map usage in virtual environments: Orientation issues. In Proceedings of Virtual Reality, pages 133–140, 1999.
12. S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In Proceedings of the International Symposium on Wearable Computers, pages 74–81, 1997.
13. A. Fischer, T. Kolbe, F. Lang, A. Cremers, W. Förstner, L. Plümer, and V. Steinhage. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. Computer Vision and Image Understanding, 72(2):185–203, 1998.
14. Google maps, 2006. http://maps.google.com/.
15. S. Güven and S. Feiner. A hypermedia authoring tool for augmented and virtual reality. In Proceedings of the International Symposium on Wearable Computers, pages 89–97, 2003.

16. T. Kim, S. Park, T. Kim, S. Jeong, and K. Kim. Semi automatic tracking of road centerlines from high resolution remote sensing data. In Proceedings of Asian Conference on Remote Sensing, 2002.
17. G. King,W. Piekarski, and B. Thomas. ARVino – outdoor augmented reality visualisation of viticulture GIS data. In Proceedings of the International Symposium on Mixed and Augmented Reality, pages 52–55, 2005.
18. C. Lin and R. Nevatia. Building detection and description from a single intensity image.Computer Vision &Image Understanding, 72, 101–121, 1998.
19. H. Mayer. Automatic object extraction from aerial imagery – a survey focusing on buildings. Computer Vision and Image Understanding, 74(2):138–149, 1999.
20. R. Nevatia and A. Huertas. Knowledge-based building detection and description. In Proceedings of the Image Understanding Workshop, pages 469–478, 1998.
21. W. Piekarski and B. Thomas. Interactive augmented reality techniques for construction at a distance of 3D geometry. In Proc. of the Workshop on Virtual Environments, pages 19–28, 2003.
22. W. Piekarski and B. Thomas. Augmented reality working planes: A foundation for action and construction at a distance. In Proceedings of the International Symposium on Mixed and AR, pages 162–171, 2004.
23. G. Reitmayr, E. Eade, and T. Drummond. Localisation and interaction for augmented maps. In Proceedings of the International Symposium on Mixed and Augmented Reality, pages 120–129, 2005.
24. G. Reitmayr, D. Schmalstieg. Collaborative AR for outdoor navigation & inform.browsing. In Proc.of Symp.onLBS&TeleCartography, pp31–41, 2004.
25. J. Rekimoto, Y. Ayatsuka, and K. Hayashi. Augment-able reality: Situated communication through physical and digital spaces. In Proceedings of the International Symposium on Wearable Computers, pages 68–75, 1998.
26. K. Satoh, M. Anabuki, H. Yamamoto, and H. Tamura. A hybrid registration method for outdoor AR. In Proc. of Internat.Symp. on AR, pp 67–76, 2001.
27. J. Shufelt. Performance evaluation and analysis of vanishing point detection techniques. Transactions on Pattern Analysis and Machine Intelligence, 21(3):282–288, 1999.
28. G. Sohn and I. Dowman. Extraction of buildings from high resolution satellite data. Automated Extraction of Man-Made Objects from Aerial and Space Images, 3:339–348, 2002.
29. G. Vosselman and J. Knecht. Road tracing by profile matching and kalman filtering. In Proceedings of Ascona Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, pages 255–264, 1995.
30. U. Weidner. Digital surface models for building extraction. Proc. of Autom. Extraction of man-made Objects from Aerial&SpaceImages,2:193–202, 1997.
31. J.Wither and T. Höllerer. Pictorial depth cues for outdoor augmented reality. In Proc. of the Internat. Symp. on Wearable Computers, pages 92–99, 2005.
32. Yahoo maps, 2006. http://maps.yahoo.com/.
33. S. You, U. Neumann, and R. Azuma. Hybrid inertial and vision tracking for AR registration. In Proceedings of Virtual Reality, pages 260–268, 1999.

Open Content Web Mapping Service: A Really Simple Syndication (RSS) Approach

Rex G. Cammack

University of Nebraska at Omaha
rcammack@mailo.unomaha.edu

Abstract

This project looks at using RSS protocols to create an environment that causes more geospatial portals users to become long-term users. RSS is a syndication method for delivering content to users. By pushing content to users on a timely basis one can address the three-stated user loss factors. For the novelty and stagnation factor providing an active content delivery system that sends timely information to the user will continually peak the user interest in the geospatial portal. Currently, most RSS aggregator software collects syndicated information solely based on time. This method works well for the content providers of aspatial information. By customizing the RSS aggregator software to collect based on both time and location the RSS protocol can work for location based services automatically.

1 Introduction

In recent years the distribution of maps has changed from a single channel of distribution into a braided channel of distribution (Fig. 1). One way of characterizing this change is by display media. Cartwright (1999) provides a detailed discussion on many of the newer multimedia technologies available to cartographers. He states that the multimedia development started in the mid-1980's with the widespread adoptions of laserdisc and CD-ROMs. A justification for the development of multimedia cartography comes from Peterson's (1999, p.35) view that paper maps were inadequate to "the essence of cartography – the representation and communication of the spatial and dynamic world". As the digital age of cartography developed, cartographers used these new media types to challenge the dominance of the paper media. Along with these new media channels, a complex set of cartographic tools was developed. Currently, cartographers must mix and match many of these new technology tools together to develop maps for multimedia, Internet and paper map products.

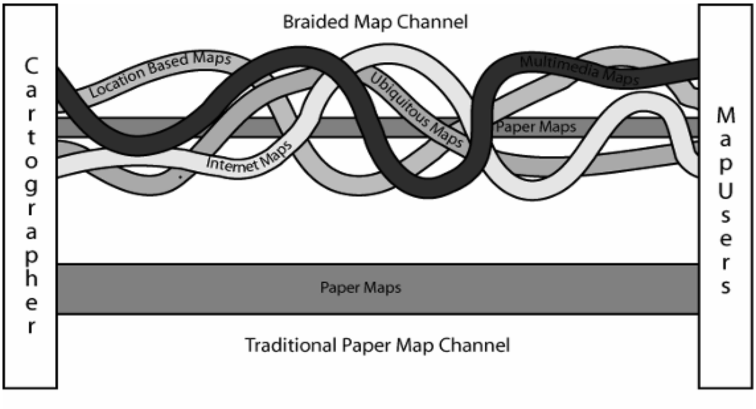


Fig. 1. Comparison of the traditional paper map and braided map channels of information exchange.

2 Portable Maps

Peterson (1999) identifies two advantages of paper maps. The first being the portability of the paper maps while the second is the better display resolution. Coupled with these advantages the paper map clearly has several disadvantages. Peterson (1999) makes a case that the distribution of paper maps has been superseded by the distribution of Internet maps. By limiting the distribution of map information to the paper channel only, spatial knowledge and understanding is held by the upper classes and/or highly educated. This control of maps reinforces the segregation of society.

Another issue to consider with making maps portable is the detachment of the cartographer from the map. Putting maps on paper and duplicating them for distribution, created a situation where the cartographer no longer accompanied the map or was allowed direct communication with the map user. This communication disconnect has long been a problem for cartography. Similarily, one of the newer trends in academics is the illustrated paper or poster session at academic conferences. The strength of this type of knowledge distribution channel is that the sender and receiver are once again in direct personal contact. Cartographers could argue that this need for direct personal contact stems from the lack of clear design of the poster. And since the majority of academics are not adequately educated in visual communication they need to rely on personal communication to clarify the message. A different perspective is that cartographers still want to accompany their maps but the distribution efficiencies have made that impossible.

Focusing on distribution efficiencies, Peterson (1999) suggested that the Internet has changed map use. Van Elzakker (2000) added that Internet maps have better accessibility and distribution. Gartner (2003) built on these ideas and coupled them with wireless technology to suggest that Internet maps can play a key role in location-based services on wireless devices. Gartner (2003) called this fusion of ideas and technology “TeleCartography”. Along with TeleCartography, the term “ubiquitous mapping” (Morita, 2005) has been used to label the combination of mapping technology, mobile display devices, wireless networks and location based services. In this research the term TeleCartography will be used. It should be noted that from the current literature the two terms have many similarities. In the future, TeleCartography and ubiquitous mapping may be demonstrated to have a clear distinction but as yet they will be treated as the same.

One could argue that the terms Internet mapping and multimedia cartography no longer have a clear distinction for cartography. The argument for this perspective would be that the map is a map regardless of the media it’s placed on. The function of the map is to store, display and express an idea between a sender and receiver. Most of the argument for the different terms comes from the technology side of the concept. Like multimedia and Internet mapping, TeleCartography involves a large amount of new technology, using the distinctive term TeleCartography seems justified.

Ormeling (1999) suggested that maps play many roles in the current Multimedia landscape including:

- Models of Spatial Reality
- Providers of Spatial Insight
- Spatial Organizers
- Tools for Accessing Information Elements

- Navigation Tools for the Multimedia Product
- Interface to the Geographic Database
- Multimedia Interface
- Vehicles for Interaction
- Interface to the Cartographic Database
- Tools for Scientific Visualization

This research takes the approach that the word TeleCartography could be substituted for the word multimedia. By using Ormeling's (1999) map roles conceptualization, a well-defined set of applications for TeleCartography has been defined.

In this introduction two primary issues have been addressed. The first one is the change in the method of map distribution from a straight paper linkage between map user and cartographer to a braided distribution path. The second issue was the identification and justification for the development of TeleCartography. The focus of this chapter will be on the linking these two issues and examining the role of syndication in map distribution and use. Maps have been a part of the syndication of information for a long time. Yet little examination of the subject has been done in regards to cartography. This research will define the concept of syndication in cartography and show how several aspects of syndication of information can help increase functionality and temporal latency of information in TeleCartography. The final point of the chapter is to show that syndication of map information can support open source mapping applications for the Internet and Location Based Services (LBS).

3 Syndication

Syndication as a term comes from the field of journalism. The idea of syndication is the process of selling the same work to numerous newspapers to be published at the same time. Many journalist/writers syndicate their work so it is published in hundreds of newspaper throughout the world. In the journalism field Ann Landers has the largest syndication with around 1200 newspapers publishing her column daily (Guinness Book of World Records, 2005). A group of people or companies that work together and produce materials used by others to build a newspaper is called a syndicate. In the publishing industry, groups like the Associated Press and Reuters are syndicates. Figure 2 shows the flow of information within the newspaper industry. It should be noted that the newspaper looks for content and selects the content that matches with their needs and style.

With this understanding of how syndicates work, this research develops a syndication framework for cartography. As discussed earlier cartographers for a long time have not been able to accompany their map to the map user. As a result of the mass production of maps, cartographers create a map product based on speculation. Companies like Michelin and Rand McNally have produced maps for years because consumers need their products. In turn, the cartographic companies place products in the market hoping that consumers will use the maps. For the consumer getting maps has been a passive activity. Map users for the most part get maps only when they either need them or when the opportunity presents itself. The majority of map users are not proactive in the acquisition of maps. The passive nature of map acquisition is the opposite of the syndication of news information.

Newspapers subscribers indicate that they want a copy before the newspaper is printed. This active approach for the distribution of newspapers has lead to the development of news syndicates. For the most part, we do not suscribe to map products. This lack of market demand for daily content has led to a passive map use community. It should be noted that one could argue that news information is fundamentally different from maps in the context of timeliness. News information is important for only a short period of time as compared to the information on the map that can be delivered later and used longer. A counter argument to the point is that some maps are timely and need to be delivered daily or in some cases more often. The best example of this timely need to deliver map information are weather maps. A weather map from three days ago does very few people much good while a weather map made ten seconds ago that shows the current conditions and projected weather over the next four hours can be used by most map users. Weather maps are already a part of most newspapers and some newspapers get their national or regional weather maps through a weather forecasting syndicate.

The idea of syndicating maps is not a novel idea in cartography or geospatial science. The concept of a web mapping service can be considered a type of map syndication. Figure 2 depicts the role of syndicates in relation to newspapers. Figure 3 shows how web mapping services fit in a syndication framework. The most important distinction between the journalism model (Figure 2) and the geospatial science model (Figure 3) is the concept of a subscription. In the journalism model the user subscribes to a service but subscribing to a syndication service is not possible for geospatial information. This may stem back from the passive nature of map acquisition for most users. The passiveness of the map acquisition might be from the lack of temporal urgency for the map user's task. The next step in making a map syndicate is developing an expectation of temporal urgency.

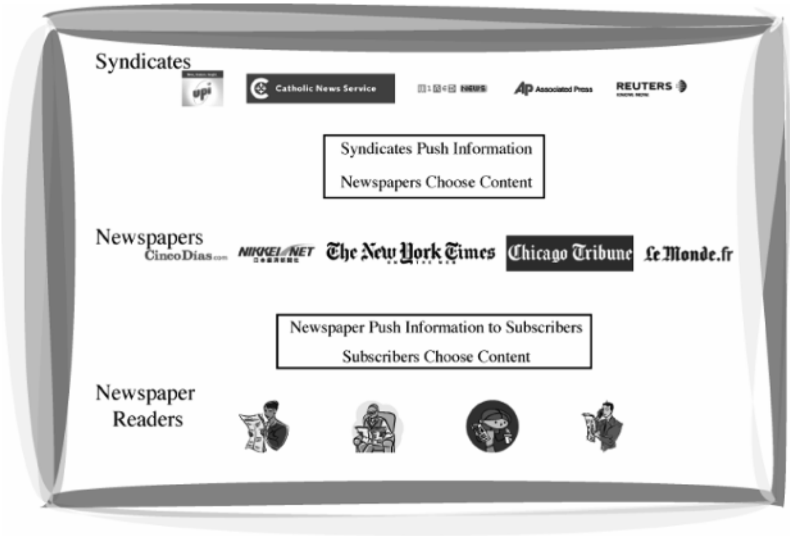


Fig. 2. How information follows through the journalism industry to newspaper readers.

Gartner (2003) has suggested that TeleCartography is an application of maps that has a temporal urgency. The integration of maps into the location based service paradigm has created an environment where maps need to be distributed to map users in a timely manner for any number of purposes. In the location based services, field map information might be available either through map brochures or publicly posted maps. But as Gartner (2003), Mount, et al., (2005), and Paelke, et al., (2005) show by using different wireless hand-held computers, maps and mapping data can be delivered to map users based on positional information of the user. In this example a subscription relation between the map user and the information provided is still missing but they all show a possible application for map subscriptions.

4 Syndication for TeleCartography

There are two possible types of maps related TeleCartography subscriptions. The first type of subscription is one that delivers maps to the map user based on some set of subscription parameters. A map subscription type of relationship would deliver maps to the map user. An important point here is the concept that the delivery of the maps is active and not passive (Figure 4). A subscription to any media is an active process. The subscriber has stated their intent to receive information based on their knowledge of the media partner before the media is developed. For the media partner, having a subscriber group allows them to tailor their products to their customer base. In the case of TeleCartography, a subscriber to a map navigation service would receive map data on demand based on both their current position or navigation destination.

The other type of TeleCartography subscription is the thematic content only subscription. An example of this type of subscription would be for Automatic Teller Machines (ATMs). One could subscribe to a bank ATM location service and every time the individual was within a set distance they could get the latest information about the ATM such as whether it is open or closed. The next section describes the technical implementation of both subscription services based on existing Internet technologies.

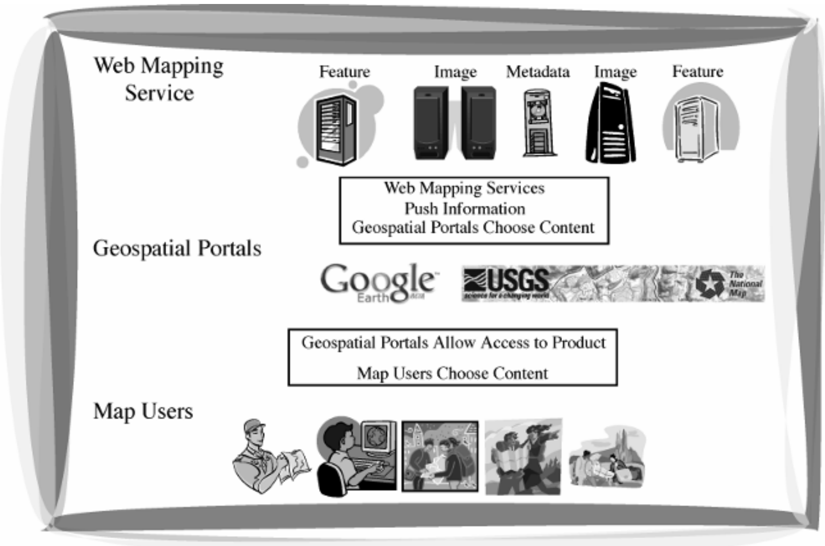


Fig. 3. How web mapping service function in a manner similar to syndicates in the journalism field.

5 Internet Syndication

The most common form of syndication on the Internet is Really Simple Syndication (RSS). The RSS standard has been around for several years and gone through several versions. Ayers and Watt (2005) give a detailed history and explanation of the difference between the different RSS standards (RSS 0.91, 0.92, 1.0 2.0, Atom 0.3 and 1.0). In recent years, the RSS standard was to be replaced by Atom because of confusion regarding the different RSS versions (see Ayers and Watt 2005 for more details). Nearly all syndicates on the Internet use the RSS or Atom standard. The process of syndication on the Internet has three parts: content provider, aggregator and user. The content provider develops the information to be distributed. The term feeds is currently being used to identify the RSS content stream. What feeds a content provider makes is up to them, but most feed content material is based on what the Internet market will consume. Currently the most popular feeds are in the field of news, opinion and music/radio. In the last 18 months the number of music/radio feeds has exploded with the development of Podcasting brought on by the popularity of the iPod™ mp3 player in the United States. As of the writing of this paper Apple computers iTunes™ website had over a thousand Podcasting sites registered with the number growing rapidly. The author has developed a Podcasting site with lecture review for his principles of cartography class (http://maps.unomaha.edu).

The second part of the syndication process is the aggregator. Numerous aggregators have been developed to gather information and then deliver the content to the user. Most of the aggregators are websites where information is aggregated at the website and tools like PHP and ASP configured pages at the request of the user. One example of this is the Google.com personal home page. Once a Google user signs into their account Google allows you to configure your page based on user preferences. Most of what Google does is allow you to subscribe to different feeds and the resulting content is arranged to your design. Other aggregators also exist inside other applications. The iTunes software from Apple computers in its 5.0 version included an audio RSS feed to handle routing. The iTunes user can select feeds from the iTunes website or any other RSS they have found. RSS feed handling is also being integrated into Internet Web Browsers. Safari has a simple RSS feed aggregator for handling any type of RSS feed.

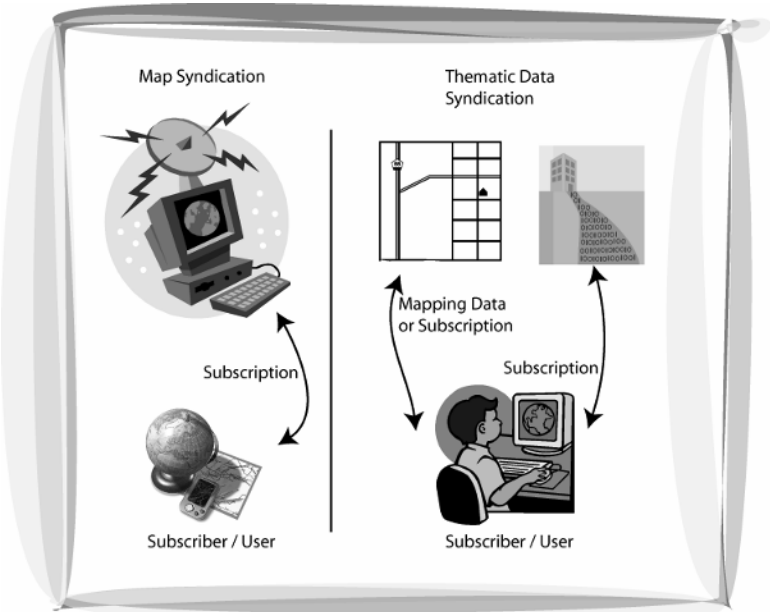


Fig. 4. The two different types of TeleCartography syndications.

The last of the three parts of Internet syndication is the user. The simplest definition of a user is any person who uses an aggregator and subscribes to any feed. The problem with this definition is the limited nature of the person. As was discussed previously in the journalism field, newspapers subscribe to news syndicates. The newspaper is much like a student who subscribes to a course lecture review podcast. So, in the context of user, the definition needs to be expanded to include any consumer of the information whether it be an end user or information redistributors.

6 RSS and Atom

RSS and Atom are built using the eXtensive Markup Language (XML). XML 1.0 was released in 1999. Since then numerous other applications of XML have been developed (Means, 2002). Figure 5 is a reproduction of an illustration from Means (2002) showing the framework of XML. Since Means' (2002) book, many other enhancements have been made to XML. One of these new developments is the Geographic Markup Language (GML). The GML standard was put forward by the Open Geospatial

Consortium and adopted in 2004 (Open Geospatial Consortium, 2004). A critical development for XML for the development of RSS was the XML Namespaces specification. Namespaces make it possible to create new markup tags that don't conflict with other tags with the same names. Many of the RSS feeds use their own XML namespaces to specify data for their aggregators. Apple Computer has developed a namespace for Podcasting feeds to work efficiently with the iTunes software (see http://phobos.apple .com/static/iTunesrss.html).

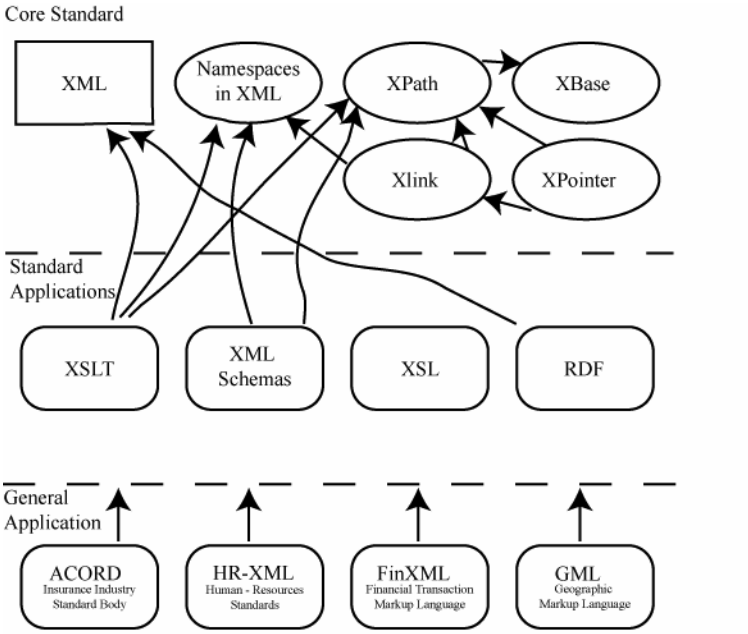


Fig. 5. The basic framework for XML and associated technologies.

Below is an example of a well-formatted XML document that has a RSS feed. The document is has four parts: XML, RSS, channel and item. The XML part of the file defines the use of the 1.0 XML specification with any standard text encoding. The RSS part of the document uses the RSS tag to define which version of RSS is used and to include any external XML namespaces developed by Apple Computer for the iTunes aggregator. The XML namespaces is one avenue to deliver maps using RSS. The third part of an RSS document is the channel. The channel tag defines a single feed. An RSS document can contain one or more feeds. In this example there is only one channel titled Cartography Podcasts. Within this channel declaration are several attributes using the iTunes: declaration. The iTunes:

declarations are defined by the podcast-1.0.dtd document included in the XML namespaces attribute. The final part of the RSS document is the item. A RSS feed can have any number of items within them. This example has only one item called “Intro to Cartography Podcast.” For most RSS feeds the item is a pointer to another file located on a website or record in a database service. One of the features of the item that has made podcasting so simple is the enclosure. With an enclosure you can send any data type associated with the item to the aggregator. So if the aggregator understands how to use the data type it will use it.

```
<?xml version="1.0" encoding="utf-8"?>
<rss xmlns:itunes="http://www.itunes.com/DTDs/Podcast-1.0.dtd" version="2.0">
<channel>
<title>Cartography Podcasts</title>
<itunes:author>Dr. Rex G. Cammack</itunes:author>
<itunes:subtitle>A show about Mapping</itunes:subtitle>
<language>en-us</language>
<copyright>&#x2117; &amp; &#xA9; 2005 Rex Cammack</copyright>
<itunes:owner>
<itunes:name>Rex G. Cammack</itunes:name>
<itunes:email>rexcammack@missouristate.edu</itunes:email>
</itunes:owner>
<link>http://maps.missouristate.edu/podcasting/cartography</link>
<description>Dr. Cammack Cartography Podcasts </description>
<itunes:category text="Science">
<itunes:category text="Cartography"/>
</itunes:category>
<item>
<title>Intro to Cartography Podcasting</title>
<itunes:author>Rex G. Cammack</itunes:author>
<itunes:subtitle>A short introduction to Cartography Podcasting</itunes:subtitle>
<enclosure url=http://utah.missouristate.edu/~rexcammack/CartCasting/welcom.mp3
        length="514704" type="audio/mpeg"/>
</item>
</channel>
</rss>
```

7 TeleCartography through RSS Feeds

In the context of Location Based Services (LBS), maps via syndication could be delivered several ways. This research examines the delivery of maps as a location based service from two perspectives.

7.1 Base Map

The first perspective is the base map. For maps to be used, a base map must be available to the LBS device. It appears that there are three ways to do this currently.

1. Precompiled and stored on the device.
2. Precompiled on servers and downloaded on demand.
3. Dynamically complied on servers and downloaded on demand.

In designing a LBS system, the method of precompiling and storing the base map on the device is the easiest to develop and test. The major drawback to the method is the need to preinstall the base maps before the user goes to the field. In most cases this approach is possible but not completely infallible. For a traveler who chooses to go to Europe with no itinerary, having the pre-loaded base map could be problematic. The second method will solve the wondering travel problem by downloading the base map based on locations when he/she needs them. This method still has the problem of pre-compilation. Cammack (2006) discusses this issue related to WMS. The synchronous development of base maps and thematic data has been an issue for cartographers for quite some time. Cammack (2005) demonstrates when either thematic data or base map data are compiled and held in a static state there is a temporal disconnect between the two parts of a map and cartographic inaccuracy will result and most likely propagate. Cammack (2006) suggests that the base map data and thematic data must be developed and maintained constantly and separately to improve the synchronous accuracy of maps.

The lack of infrastructure is a major impediment for TeleCartography for both methods two and three. In areas with low population density the likelihood that public wireless networking will be developed is improbable. A historical example of this point is the Tennessee Valley Authority (TVA). In sections of the southern part of the USA electrical networks would have never been built without direct federal government intervention. The manner in which publicly accessible wireless networks are being created in the USA, it will result in low population areas most likely lagging behind areas of higher population density. This is not an unknown factor in LBS and some hope may rest with a space-based solution to this problem. Although this drawback exists it should not deter the development of TeleCartography for LBS.

To deliver base maps for LBS using RSS requires several steps. The first is to develop an RSS feed similar to the example above. Instead of delivering MP3 files the feed would enclose an Internet standard

MIME/TYPE that is a graphic format. The graphic format would need to be consistent with the TeleCartography LBS application. A key aspect to developing this type of RSS feed is to create standard XML namespaces for this type of RSS feed and enclosure. Apple (2005) created XML namespaces for iTunes and a similar standard could be developed for RSS feeds for TeleCartography and LBS. The XML namespaces would need to create attributes for RSS enclosures that could be used by the LBS device to select the correct RSS enclosure to display. One must consider what triggering mechanism will be used to get a map for the LBS device. Triggering mechanisms for LBS tend to be locational which is a natural fit for TeleCartography and LBS. With the development of a RSS feed and a triggering mechanism, the next task is to create an aggregator that will handle the RSS feed subscription and understand the new XML namespaces and MIME/TYPE in which the map is delivered. Currently the XML namespaces and aggregator does not exist but development in both areas is underway.

7.2 Thematic Data

The second aspect of TeleCartography for LBS using RSS is the delivery of current thematic data. Currently most LBS are fee-based service. Cost as limiting factor and needs to be addressed. Some LBS charge users by either a monthly fees or per use charges. There is probably no way to create a completely free TeleCartography LBS so cartography must develop a means of controlling costs. One way to control cost is to make access to maps less restricted. The section on base maps shows a possible way to make an open system using RSS that would work with many TeleCartography providers. A second way the LBS provider makes money is by selling their services to point of delivery business. A point of delivery business is a business that has a store that a customer enters and makes a financial transaction, i.e., bank, supermarket, drugstore or beer hall.

The LBS provider charges fees to the LBS user and the point of delivery business. If a point of delivery business chooses not to pay this service then the user is limited by not knowing about the business. Consider this example, an LBS user is in Vienna and hungry for pizza but doesn't know where the pizza businesses are located. Being an LBS user they call up the map based navigational tool and ask for pizza stores. The LBS provider had marketed this service to all the pizza business and some have chosen to pay to advertise on this LBS service. The consumer will only be able to choose from the restricted list. If we take this example one step further and the LBS user knows that they want to go to Pizza Mann but Pizza Mann has chosen not to advertise on this LBS. The LBS provider has failed to provide the user with his desired information.

RSS feeds could address this problem. By making a TeleCartography system that is RSS savvy, the user could subscribe to any geolocational RSS feed. The system would work like this. First, the point of business would create a RSS feed for store location and information. Within the RSS feed, each store location would be enter attributes. What makes this a geolocational RSS feed is the inclusion of the Basic Geo vocabulary (W3C, 2003).

The W3C Semantic Web Interest Group developed the Basic Geo vocabulary. The result of this effort was the creation of a basic geo RDF vocabulary for RDF documents. A second outcome was the use of GEO XML namespaces that is used with RSS feeds. Yahoo! Maps (Yahoo!, 2005) used the GEO XML namespaces to interact with the Yahoo! Maps service.

For the cartographer making a TeleCartography or LBS system, making systems that are RSS savvy and have the ability to geocode simple latitude and longitude coordinates would allow the user to subscribe to any geolocational RSS service and have access to their information in a spatial format. This real-time thematic mapping would eliminate some of the limitations of some of the current LBS systems.

8 Conclusions

This research has examined the role of syndication of maps for TeleCartography and LBS. The ability to syndicate maps across the Internet is a fundamental change in the information exchange process that cartographers have used for hundreds of years. The syndication process allows the cartographer to have some input from the map user before maps are actually made and distributed. A second point to the research was a detailed examination of existing technologies to see if the tools are currently available to map syndication possible for LBS map syndication. The results show that it would be possible to syndicate maps quickly and simply by using RSS feeds and developing an image MIME/TYPE aggregator. In addition, thematic data could be syndicated in a geolocational RSS feed using the GEO XML namespaces. By adding a simple geocode function to a RSS aggregator, a TeleCartography and LBS system could become more accurate for temporal information.

References

Apple (2005). How to Publish a Podcast on the iTunes Music Store. http://phobos.apple.com/static/iTunesrss.html last modified 2005-07-07.

Ayers, D. and A. Watts (2005). Beginning RSS and Atom Programming. Indianapolis USA: Wiley.

Cammack, R. G. (2005). Web Mapping Services: A Tool For Thematic Internet Maps. In the Proceedings of XXII International Cartographic Conference (ICC2005). A Courna, Spain.

Cammack, R. G. (forthcoming Publication). Cartography Approaches to Web Mapping Services. Multimedia Cartography. 2nd ed Berlin, Germany: Springer.

Cartwright W. (1999) Development of multimedia. In Cartwright W., M. Peterson and G. Gartner (eds). Multimedia Cartography. Berlin, Germany: Springer.

Peterson, M. (1999). Elements of Multimedia Cartography. In Cartwright W., M. Peterson and G. Gartner (eds). Multimedia Cartography. Berlin, Germany: Springer.

Gartner (2003). TeleCartography: Maps, Multimedia and the Mobile Internet. In M. P. Peterson (eds). Maps and the Internet. Oxford, UK: Elsevier.

Guinness World Records (2005). Guinness World Records 2006. Lincolnshire, UK: Guinness.

Means, W. S. (2002). Strategic XML. Indianapolis, USA: Sams Publishing.

Morita, T. (2005). A Working Conceptual Framework for Ubiquitious Mapping. In the Proceedings of XXII International Cartographic Conference (ICC2005). A Courna, Spain.

Mount J., M Armstrong, and D. Bennet (2005) Field-Based Geographic Education Using Mobile Computers. Internet-Based Cartographic Teaching and Learning: Atlases, Map Use and Visual Analytics. Madrid, Spain International Cartographic Association.

Open Geospatial Consortium (2005). Geographic Markup Language Encoding specification. http://www.opengeospatial.org/specs/?page=baseline.

Ormeling, F. (1999). Map Concepts in Multimedia Products. In Cartwright W., M. Peterson and G. Gartner (eds). Multimedia Cartography. Berlin, Germany: Springer.

Paelke, V., B. Elias, and M. Hampe (2005). The CityInfo Pedestrian Information System: An Experiment with Content Creation and Presentation Techniques. In the Proceedings of XXII International Cartographic Conference (ICC2005). A Courna, Spain.

Van Elzakker, C. (2000). Users of Maps on the Web. In Kraak, M. and A. Brown (eds) WebCartography. London, UK: Taylor and Francis.

W3C (2003). W3c Semantic Web Interest Group: Basic Geo (WGS84 lat/long) Vocabulary. http://www.w3.org/2003/01/geo/#development .

Yahoo! (2005). Introduction to Yahoo! Map Web Service. http://developer.yahoo.net/maps/.

Section V Applications

Chapter 31	Carlo Ratti, Andres Sevtsuk, Sonya Huang, Rudolf Pailer Mobile Landscapes: Graz in Real Time
Chapter 32	Rein Ahas, Jaak Laineste, Anto Aasa, Ülar Mark The Spatial Accuracy of Mobile Positioning: Some experiences with Geographical Studies in Estonia
Chapter 33	Natalia Andrienko, Gennady Andrienko A Framework for Decision-Centred Visualisation in Civil Crisis Management
Chapter 34	Erich Wilmersdorf Providing an Information Infrastructure for Map Based LBS - The approach of the City of Vienna
Chapter 35	David A. Bennett, Marc P. Armstrong, Jerry D. Mount MoGeo: A location-based educational service
Chapter 36	B.J. Köbben, D. Boekestein, S.P. Ekkebus, P.G. Uithol Bata Positioning System - A real time tracking system for the world's largest relay race
Chapter 37	Karl Rehrl, Nicolas Göll, Sven Leitinger, Stefan Bruntsch, Hans-Jörg Mentz Smartphone-based information and navigation aids for public transport travellers
Chapter 38	Lilian Pun-Cheng, Esmond Mok, Geoffrey Shea, W.Y. Yan EASYGO – A public transport query and guiding LBS
Chapter 39	Siegfried Wiesenhofer, Helmut Feiertag, Markus Ray, Lucas Paletta, Patrick Luley, Alexander Almer, Mathias Schardt, Josef Ringert, Paul Beyer Mobile City Explorer: An innovative GPS and Camera Phone Based Travel Assistant for City Tourists
Chapter 40	Byoung-Jun Kang, Yosoon Choi, Hyeong-Dong Park Development of Cultural Inheritance Information System using LBS Technologies for Tourists
Chapter 41	Michaela Kinberger, Karel Kriz, Patrick Nairz LWD-Infosystem Tirol – visual information about the current avalanche situation via mobile devices
Chapter 42	László Zentai, Antal Guszlev Spatial tracking in sport

Mobile Landscapes: Graz in Real Time

Carlo Ratti[1], Andres Sevtsuk[1], Sonya Huang[1], Rudolf Pailer[2]

[1] SENSEable City Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
[2] mobilkom Austria, Vienna, Austria
ratti@mit.edu

Abstract

The technology for determining the geographic location of cellphones and other hand-held devices is becoming increasingly available. It is opening the way to a wide range of applications, collectively referred to as Location Based Services (LBS), that are primarily aimed at individual users. However, if deployed to retrieve aggregated data in cities, LBS could become a powerful tool for urban cartography. This paper describes preliminary results of the "Mobile Landscapes: Graz in Real Time" project, which was developed as part of the M-City exhibition (Graz Kunsthaus, 1 October 2005 – 8 January 2006, curator Marco De Michelis), in collaboration with the cellphone operator mobilkom austria. Three types of maps of the urban area of Graz, Austria, were developed and shown in real-time on the exhibition premises: cellphone traffic intensity, traffic migration (handovers) and traces of registered users as they move through the city. Beyond their novelty and visual interest, results seem to open the way to a new paradigm in urban planning: that of the real-time city.

1 Introduction

The mobile communications industry is booming. Cellphone subscriptions have recorded sustained growth rates in recent years and, according to EITO (2004), reached the astronomic figure of 350 million in Western Europe in 2003 (157 million in the USA).

Why should the cartography and urban planning community be interested in this data? First, the widespread deployment of mobile communications, supported by personal handheld electronics, is having a significant impact on urban life. People are changing their social and working habits because of the new technology (Rheingold, 2002). Activities that once required a fixed location and connection can now be achieved with higher flexibility, resulting in the users' ability to act and move more freely (for an analysis in the corporate working domain, see Duffy, 1997). As a consequence, urban dynamics are becoming more complex and require new analysis techniques. Second, and more importantly in this context, data based on the location of mobile devices could potentially become one of the most exciting new sources of information for urban analysis.

Locational data are becoming increasingly available and their applications are currently a hot topic in the cellphone industry (see for instance www.lbszone.com). They are generally referred to as Location Based Services (LBS) – value-added services for individuals in the form of new utilities embedded in their personal devices. Examples, both implemented and speculative, include systems providing information about one's surroundings (nearby restaurants, museums, emergency shelters, and so on); distributed chat lines aimed at allowing people with similar profiles to encounter each other in space, via a kind of technologically augmented serendipity; and 'digital tapestries' that attach different types of information to physical spaces (see sections below for detailed references). And yet, surprisingly enough, aggregated locational data have not been used to describe urban systems. Research efforts in the area are sparse; the scientific literature mostly ignores themes such as the mapping of the cellphone activity in cities or the visualization of urban metabolism based on handset movements (notable exceptions are the work of Ahas and Ülar, 2005 and Ratti *et al.*, 2005). How could this be?

We will try to guess. The first assumption is that scholarly research has been hampered so far by the difficulty of accessing raw data and developing ad hoc analysis software and systems in partnership with cellphone companies. The second reason could be traced back to the lack of rules for managing these data and to the increasing privacy concerns that are being raised, often resulting into stalling actions. In this study, the research team has had the opportunity to establish a partnership with the mobile network operator mobilkom austria, which has the largest market share in its own country. Thus, a privileged insight into how aggregated data from mobile devices could reveal urban systems was gained. Furthermore, the occasion of the project was developed in the context of a public architecture and art event, the M-City exhibition (Graz Kunsthaus, October 2005 – January 2006, curator Marco De Michelis). Such a premise seemed ideal in order to remove suspicions of possible privacy abuse and openly engage the public in the issues related to locational data and the way they should be used.

Only some preliminary results are presented here. Further analyses will be performed in the coming months. However, results seem to open the way to a new approach to the understanding of urban systems, which we have termed “Mobile Landscapes”. Mobile Landscapes could give new answers to long-standing questions in architecture and urban planning: How to map vehicle origins and destinations? How to understand the patterns of pedestrian movement? How to highlight critical points in the urban infrastructure? What is the relationship between urban forms and flows? And so on. More generally, information and communication technologies, which previously eluded planners with their “invisible, silent” characteristics (Graham and Marvin, 1996), are becoming of increasing interest. Studies on user behavior (for example Charlot and Duranton, forthcoming; Hampton and Wellman, 2000; Gaspar and Glaeser, 1996) are showing that “electronic communications” and the “metropolitan area […] are actually supporting each other” (Graham, 2004). The aim here is to expand upon these methods by integrating a spatial component to visualize telecommunications. Thus, the objectives of our maps are twofold: first, they enable viewers to visualize the city via an otherwise “invisible” feature; second, the maps are animated in a way that can simulate city dynamics in real time.

In a previous paper (Ratti *et al.*, 2005) we have shown how even simple data such as network traffic have a lot to offer to the urban designer or planner. We will describe here how some more complex types of data can be obtained to produce real-time urban maps. In particular, three types of data are provided by the mobile network, namely cellphone traffic intensity, traffic migration (handovers) and traces of registered users as they move through the city.

2 The data

Mobile communication networks are organized in cells. Each cell has an antenna that covers a certain geographical area (usually circular or pie-shaped) and offers mobile communication services to users at that particular location. In the most common GSM networks, cell sizes vary from the diameter of about 100-300 meters in urban areas to several kilometers in rural areas. Densely inhabited areas require smaller cell sizes, as the maximum amount of data that can be transferred concurrently in one cell is limited and relates to the radio access technology of the network.

Fig. 1 shows the basic components of a GSM network and the data flow that was implemented for the Mobile Landscape Graz project. A Mobile Switching Center (MSC) is connected to several Base Station Controllers (BSC). The BSCs control the Base Transceiver Stations (BTS). The Base Transceiver Station houses the radio transceivers that define a cell and handles the radio-link protocols with the Mobile Equipment (ME). The movement of a mobile phone from one cell to another is called ‘handover’.

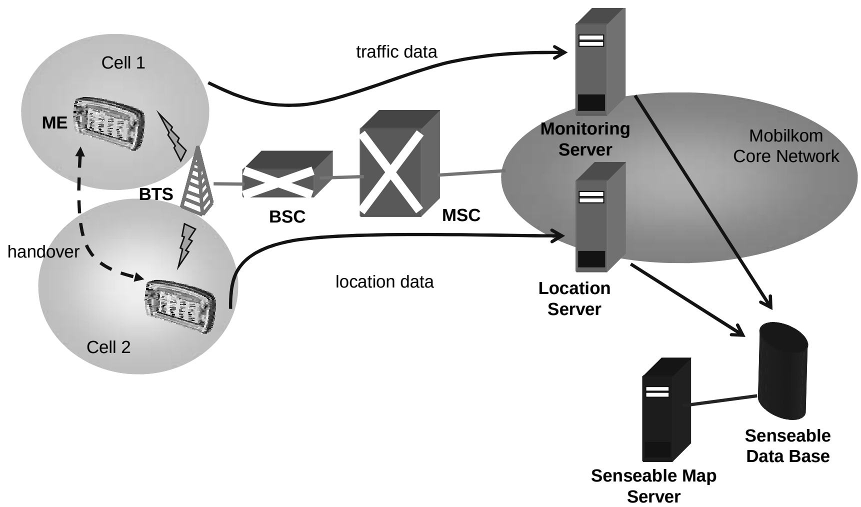


Fig. 1. Data flow in the project M-City

Traffic in the cells of the mobilkom austria network is constantly monitored by the Operation and Maintenance division and aggregated performance figures are logged by a Monitoring Server. For the Mobile Landscape Graz project the traffic of all cells in the city of Graz was measured periodically and traffic aggregates were reported at regular intervals to an online database set up by the MIT SENSEable City Laboratory. Traffic intensity of one cell is measured by the number of calls that originate there and by the occupancy of the cell measured in Erlang, a standard unit of measurement of traffic intensity in a telecommunications system (one Erlang is the equivalent of one caller talking for one hour on one telephone or two callers talking for 30 minutes each). Table 1 shows a sample of traffic intensity data, where BSC is the name of the Base Station Controller, FZA is an identification of the Base Transceiver Station, CI is the cell ID, TCH_ATT gives the number of traffic channel attachments including handovers, Erlang gives the occupancy of the cell and CALL_REQ gives the number of originated calls.

Table 1. Traffic intensity data sample

int_id	Period_start_time	bsc	fza	ci	tch_att	erlang	Call_req
792903	2005-10-08 05:30:00	Graz_4	G285-1	28516	8	0.13	4
793754	2005-10-08 06:30:00	Graz_4	G285-1	28516	8	0.40	3
794605	2005-10-08 07:30:00	Graz_4	G285-1	28516	38	0.89	17

Traffic migration is measured by the number of active calls that move from one cell to another and is reported to the SENSEable City Laboratory by giving the number of incoming and outgoing handovers between cell pairs. Table 2 shows a sample of the traffic migration data where the number of inbound and outbound handovers (HO_in, HO_out) are given for source and destination cells (Source_CI, Dest_CI).

Table 2. Traffic migration sample data

int_id	Period_start_t	source_bsc	source_fza	source_ci	ho_out	ho_in	dest_ci
7089747	2005-10-10 10:00:00	Graz_1	G200-2	20026	3	4	58236
7089746	2005-10-10 10:00:00	Graz_1	G200-2	20026	2	13	43130
7089745	2005-10-10 10:00:00	Graz_1	G200-2	20026	10	25	43110

Traffic intensity and traffic migration are aggregate figures that show in real time where people started their mobile phone calls, how long they talked and where they were moving while talking. The third type of data that was shown on the maps at the M-City exhibition was the movement of individual users. While the measured traffic aggregates collect only data about active phone calls, a cellular network is also able to locate passive mobiles in order to set-up calls towards them. This process is called ‘paging’ and can also be used to request the ID of the cell to which the mobile phone is attached, even if no call is being made. Based on the cell ID, geographical coordinates can be obtained, corresponding to the central point of the GSM cell where the user is attached to the network. Movement tracks can thus be generated by locating the user at regular 5-minute intervals.

A note should be made, however, on the treatment of location data. On the one hand they can be of considerable value to information and communication services. On the other hand, however, they raise privacy issues; users are concerned about revealing their position data to others, especially to un-trusted third party applications. Furthermore, most countries have legal restrictions that regulate processing of personal data and the protection of privacy in electronic communications. It is of utmost importance that the users can control who gets access to their location data and that the transport in the network of such sensitive data is protected by strong security mechanisms.

The Mobile Landscape Graz installation pays respect to privacy concerns by giving users full control over the location service. Only volunteers are tracked and they have to subscribe actively to the service by sending a short message (SMS), where they can also provide a pseudonym (nickname) that is shown on the map. If no pseudonym is provided, a random one is generated. The user can stop the location application at any time by simply sending a second SMS to the subscription application. Even if no second SMS is received, the application stops locating users after 24 hours. Location data is sent to the SENSEable City Laboratory via an encrypted connection and it contains only users' pseudonyms. Table 3 shows a data sample where the geographical coordinates in latitude and longitude are given for several pseudonyms at regular intervals.

Table 3. User movement sample data

nick	latitude	longitude	timestamp
Sabine	47.06354	15.454447	20051009-14:32
User940	47.06354	15.454447	20051009-14:32
Wort	0.0	0.0	20051009-14:32
Nyko	47.05407	15.462827	20051009-14:32
OrangeMo	47.10022	15.397057	20051009-14:32
User3346	47.09155	15.7262	20051009-14:32

3 The real-time mapping system

A common procedure has been developed for all data types presented above (cellphone traffic intensity, traffic migration and traces of registered users as they move through the city) in order to produce real-time maps at the exhibition premises. Records were collected live by mobilkom austria and sent to an ad-hoc set up server at MIT, where they were stored in an open-source MySQL database. The data was then paired using ArcGIS with spatial information on the city of Graz, generously provided by GIS Steiermark (www.gis.steiermark.at). Finally, the whole process was animated using Macromedia Flash. A diagram of the process is shown in figure 2.

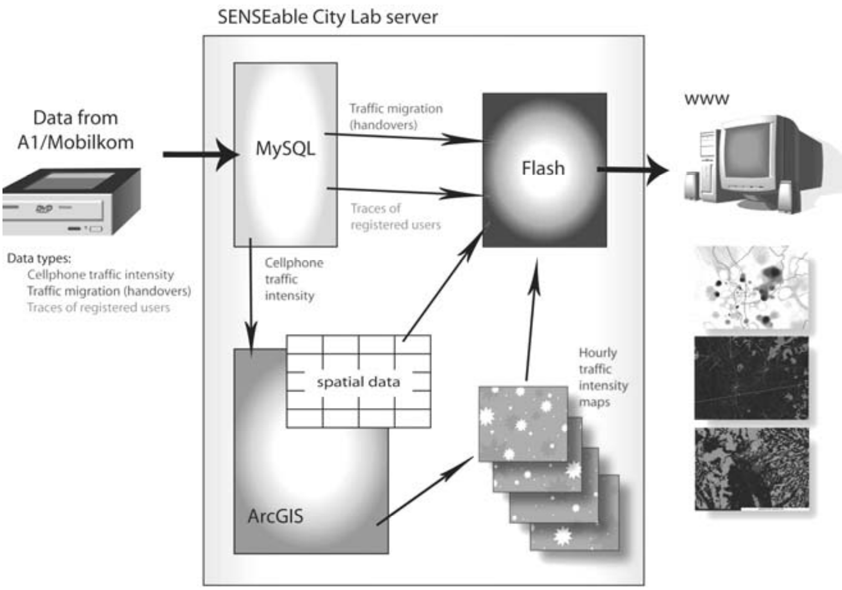


Fig. 2. Data flow from the SENSEable City Laboratory server to the final maps

A note should be made on the procedure and on the choice of using Flash. Streaming real-time animations from the MIT SENSEable City Laboratory to the exhibition venue would have been too costly in terms of bandwidth and also unreliable. Flash allowed high compatibility interface, running on most Internet browsers, and also the transfer of a limited amount of data. In the case of the Erlang map, just a maximum of 24 images were transferred from SENSEable City Laboratory and then animated in Flash. In the case of the two other maps, a Flash algorithm running at the exhibition venue retrieved automatically the data from the MySQL database to create the required animations, as detailed below.

Options other than Flash were tested, but did not prove satisfactory. Adopting ArcGIS to produce the animations would have required end users to have a GIS viewer – thus limiting the distribution of the images. Options for producing maps on the web without end-user software were also considered, such as Mapserver and ArcIMS; however, both technologies did not have the ability to reproduce the real-time feel of animated images and lacked the graphic qualities we were aiming for.

4 The results

This paper presents work in progress, as a larger interpretative effort is still underway. However, this section of the paper presents the maps that are being produced and discusses their meaning.

Using the three types of data transferred by mobilkom austria to the SENSEable City Laboratory, three different real-time maps of Graz were obtained: cellphone traffic intensity, traffic migration (handovers) and traces of registered users as they move through the city. These maps visualize different states of cellphone activity in Graz and the emergent movement patterns of cellphone users throughout a day. Our aims were manyfold: first, by cartographically illustrating the raw cellphone traffic data, we were trying to visualize dynamic layers of information, which tend to remain invisible in everyday life. Second, we were trying to use cellphone information as a substitute for other urban information, thus visualizing the city as a real-time, pulsating entity. Third, we were trying to display different types of sensitive information, such as location data, in order to engage the public in the discussion on how it should be treated in the coming years. More details follow on the three different types of maps, presented in figure 3.

4.1 Cellphone traffic intensity

The Erlang map shows the most recent distribution of cellphone traffic intensity in Graz. It is animated in such a way to play through a three-minute visualization of the last 24 hours. The data shows total Erlang values of each mobilkom austria cell antenna, interpolated using the color-field intensity map shown in the lower right corner of the image. Light and pink colors stand for higher intensity, and green and black for lower or zero intensity respectively. A street and river map of Graz is overlaid on top of the interpolated graph in order to provide location references for exhibition visitors.

4.2 Traffic migration

Similar to the Erlang maps, data is received at hourly installments and then animated in an accelerated way over 15-minute cycles. For each handover, a dynamic orange line is drawn using Flash between its origin and destination. If several handovers happen between the same pair of cells, several lines are drawn over the same trajectory. Already drawn lines slowly fade away in order to make new ones more visible. The animation renders all the phone-calls of a randomly chosen cell in a sequence and then jumps to another cell. This is why star-like shapes appear on the map - a star representing the total incoming and outgoing calls of a given cell. If the origin and destination of a handover lie outside the visible area of the map, then the call is still mapped as a straight orange line. At the end of a 15-minute cycle, the total activity graph of the last hour becomes visible.

4.3 Traces of registered users

This map shows individual mobilkom austria clients who registered to have their cellphone tracked on the map. The registration process happens by sending an SMS to an activation number. From that moment onwards, the selected nickname of the user appears on the map and her/his position is followed at five-minute intervals. A user can also stop being tracked anytime, by sending a “stop” message to the same number; alternatively, he/she will be automatically withdrawn after 24 hours. Traces are visualized as orange lines on the map, showing movements during the past 24 hours. In order to be able to differentiate between different paths on a complex background generated by multiple users, the following procedure was implemented: one after the other, paths are highlighted in red and scanned by a white ball which replays the past 24 hours, while the nickname of the corresponding user is shown on the lower left corner of the screen.

Fig. 3. From left to right: visualization of cellphone traffic intensity on 10 October 2005 at 8 pm; visualization of traffic migration (handovers) at the same time; traces of registered users at the same time.

5 Discussion

The maps produced in partnership between mobilkom austria and the SENSEable City Laboratory proved to be successful in visualizing the city of Graz in real time. This proof of concept seemed quite important, as no similar precedents based on cellphone mapping are found in the scientific literature.

The work also managed to capture wide interest at the exhibition in the Kunsthaus Graz as well as on-line and in the press (some of the published articles and discussion can be found at the URL: http://senseable.mit.edu). Visitors in Graz seemed highly interested in seeing their home-town represented in a new, intangible way and were keen on testing the tracking system to follow their own or their friends' traces. Many reviews on the project also confirmed both the lack and necessity of such cartography to add to our urban knowledge.

The public appearance of the project also brought up many discussions about privacy concerns. Even if they were not so extreme as the 'Geoslavery' fears voiced by Dobson and Fisher (2003), before the official opening of the exhibition 'big brotherish' comments appeared amongst participants and in the media. However, as privacy procedures were thoroughly respected at every step of the project and individual tracking could only happen on a voluntary basis, none of these points became detrimental. Conversely, they contributed in a certain sense to one of the aims of the project: presenting this new type of sensitive information in order to prompt a discussion on how it should be used.

In a previous paper (Ratti *et al.*, 2005), we have shown that even simple cellphone traffic analysis can make an exceptional contribution to urban analyses. More generally, there seems to be a large gamut of interests, ranging from the enrichment of people's understanding about urban communications to the unprecedented collection of real-time data about cities (see William Mitchell's (2005) review of the Mobile Landscape Graz project, "The Real Time City").

Finally, it is worth mentioning some limitations of our process. First, as mentioned above, the accuracy of phone call and user locations is estimated using the latitude and longitude coordinates of cells. This means that the larger the diameter of the cell, the less accurately we can determine the user's exact position. Furthermore, users who are roughly equidistant between two cell antennae may have their signals picked up by either of them, sometimes even bouncing between cells. The bouncing phenomenon can be observed on our traces map: sometimes static users appear to move back and forth repeatedly between neighboring antennae.

Finally, the Flash software lacks ArcGIS's precision to convert geographical coordinates to pixel-based screen coordinates. Although we kept data in ArcGIS for as much of the procedure as possible, the final representation in Flash introduces some distortion in the mapping. This was acceptable in the context of the exhibition, but might require a different treatment in future quantitative analyses.

A last note on the notion of 'real time': while data were collected instantly by the network operator, their transfer to the SENSEable City Lab server was aggregated at 5-minute to 1-hour intervals. This was done to optimize bandwith and processing power, though the same system that was set up could theoretically withstand second-long (or less) frequencies.

6 Conclusions

This paper reports on the Mobile Landscape Graz project, developed by the MIT SENSEable City Laboratory in collaboration with mobilkom austria as part of the M-City exhibition (Graz Kunsthaus, 1 October 2005 – 8 January 2006, curator Marco De Michelis). The different types of data, the procedure set up to transfer them in real time and the resulting maps are described above. While a larger interpretative effort is planned, preliminary results suggest that the mapping of cellphone data could open unprecedented perspectives in urban cartography and lead to a new urban paradigm: that of the real time city.

Acknowledgements

First, we would like to thank all of the SENSEable Graz team: Daniel Berry, Daniel Gutierrez, David Lee, Xiongjiu Liao, Jia Lou, Andrea

Mattiello, Eugenio Morello. They are the ones who made this project happen! Acknowledgements go to ‘M-City’ curator Marco de Michelis and all of the people at the Graz Kunsthaus (in particular, Peter Pakesch, Sabine Suppan and Katrin Bucher). Also, the project would not have been possible without the generous support of mobilkom austria, who provided us with the data and worked closely with our team. Their continuing involvement proves their farsighted commitment to academic discovery and policy-making. Finally, the results reported in this paper are part of a broader research effort on the use of new technologies to describe cities. We are indebted to many people at the Massachusetts Institute of Technology, Cambridge, MA, for providing an extremely stimulating research environment and for their generous feedback; amongst others: Eran Ben-Joseph, Joseph Ferreira, Dennis Frenchman, Hiroshi Ishii, Frank Levy, William Mitchell, George Stiny, and Lawrence Vale. Janet Owers, based in the ‘real’ Cambridge (UK), provided expert editorial guidance. Of course, any shortcomings are our sole responsibility.

References

Ahas R., Ülar M., 2005, “Location based services—new challenges for planning and public administration?”, Futures, 37:6 547-561

Charlot S., Duranton G., 2005, “Cities and workplace Communication: Some French Evidence”, Urban Studies (forthcoming)

Dobson J, Fisher P, 2003, Geoslavery, IEEE Technology & Society 22 (1), 47-53

Duffy F., 1997, The New Office, Conran Octopus, (London)

EITO, 2004, Report by the European Information Technology Observatory EITO http://www.eito.com/index-eito.html

Gaspar J., Glaeser E.L., 1996, Information Technology and the Future of Cities, NBER Working Paper No. W5562 (May 1996)

Graham S., 2004, "Beyond the Dazzling Light: From Dreams of Transcendence to the 'Remediation' of Urban Life", New Media and Society, 6 33-42.

Graham S., Marvin S., 1996, Telecommunications and the City: Electronic Spaces, Urban Places, Routledge (London)

Hampton K.N., Wellman B., 2000, “Examining Community in the Digital Neighborhood: Early Results from Canada's Wired Suburb”, Lecture Notes in Computer Science, 1765 194-208

Mitchell W., 2005, “The Real Time City”, Domus, October 2005, 885 94-99

Ratti C., Pulselli R. M., Williams S., Frenchman D., 2005, “Mobile Landscapes”, Environment and Planning B – Planning and Design, forthcoming.

Rheingold H., 2002, Smart Mobs: The Next Social Revolution, Perseus Books Group (Cambridge, MA)

The Spatial Accuracy of Mobile Positioning: Some experiences with Geographical Studies in Estonia

Rein Ahas[1], Jaak Laineste[2], Anto Aasa[1], Ülar Mark[3]

[1] Institute of Geography, University of Tartu, Vanemuise St 46, Tartu 51014, Estonia, tel: +372-5035914, fax +372-7375825, rein.ahas@ut.ee
[2] Regio Ltd., Riia 24, Tartu 51010, Estonia, jaak.laineste@regio.ee
[3] Positium ICT, Niine 11, Tallinn, Estonia, ular@positium.ee

Abstract

This paper describes some aspects of mobile positioning accuracy when using mobile positioning data in geographical studies. The method used in Estonia is named the Social Positioning Method (SPM) and uses locations of mobile phones and the personal characteristics of phone owners for studying human behaviour. Positioning experiments conducted in Tallinn and Tartu since 2003 have shown that SPM data facilitates the successful analysis of the space-time behaviour of society. The calculations of theoretical positioning error based on 180 000 positioning measurements in the Estonian GSM network (CGI+TA positioning method) in 2004 showed that 61 percent of positioning points are accurate to within 1000 meters in urban areas and 53 percent are accurate to within 3000 meters in rural areas. Accuracy checks conducted using GPS showed that 52 percent of positioning points are accurate to within 400 meters in urban areas and 50 percent are accurate to within 2600 meters in rural areas. While some of the research findings are limited because of the low accuracy of the positioning data, many research directions, which use a smaller scale such as commuting and regional development studies are nevertheless very promising.

1 Introduction

There have been many discussions about applications and businesses involving mobile positioning and location based services (LBS) but only a few of these have become widespread. One important reason for the lack of revolutionary developments in LBS applications is the limited accuracy of positioning (Mannings et al 2003; Spinney 2003; Ahas and Mark 2005). North American and Japanese network standards allow the use of GPS assisted (A-GPS) telephones today, this technology is developing also to European GSM networks. For example, A-GPS will start operation in end of 2006 in the Estonian EMT network. Today's available mobile positioning information demonstrates a wide range of positioning methods and accuracy, which may differ from ten meters to tens of kilometres (Ahonen and Eskelinen 2003; Adams et al 2003). Accuracy has actually been measured in many networks but not much data has been published mostly due to business interests and the rapid development of the technology and methodology of positioning (Laineste 2003).

This paper presents some results of the measured accuracy of mobile positioning in the context of geographical and social studies. The data and experiences obtained during the experiments of the social positioning method (SPM) have been carried out in Estonia since 2003. SPM uses the movement co-ordinates of mobile phones and the characteristics of their owners in studying the space-time behaviour of society and has been developed by the Institute of Geography, University of Tartu and Positium Ltd in Estonia (Ahas et al 2006; Ahas and Mark 2005). The application of location based services has been used and described by geographers in many institutions (Mountain and Raper 2001; Spinney 2003; Asakura and Hato 2004; Ohmori et al 2005). There have been several studies carried out using passive mobile positioning data as data of network use (load) by unidentified phones. Still relatively few studies have applied positioning in research projects with personal identities as there are problems with personal privacy and surveillance fear.

2 Mobile positioning and data collection

2.1 Positioning methods

Mobile positioning data used in geographical studies can be divided into passive and active. *Passive mobile positioning data* - location data which is stored automatically in some kind of log files (as billing or handover memory files) of mobile operators whenever a person's mobile phone interacts with the mobile network (typically during call or message sending/receiving activity) (Ahas et al 2006). Passive positioning data can be recorded specifically for call activities or as a measure of the load of phones in network cells. *Active mobile positioning data* - mobile tracing data where the location of a mobile phone is determined (asked for) using a special query or a request using radio waves. These active location requests can be phone- or network initiated. This type of data is used in SPM studies in Estonia presented in this paper.

Technical solutions for tracing mobile phone location actively are different for different network standards and hardware producers (Adams et al 2003; Ahas and Mark 2005). These positioning solutions can be handset based, network based, or satellite (mostly GPS) based. Network based methods use different positioning techniques to find phone locations in a network: cell identity; triangulation; angle; arrival time of radio signal etc. The most commonly used methods use the unique ID of individual network cells: cell global identity (CGI). Network cells can be relatively large, range between 0.002 to 100 km2.

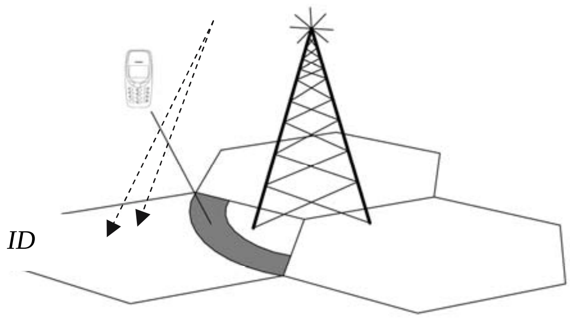


Fig. 1. Cell Global Identity + Timing Advance (CGI+TA) method for mobile positioning, used in SPM studies in Estonian EMT network with Ericsson positioning server.

As the network cells are too big for determining the exact location of a phone, the exact location (distance) in a cell is determined by timing advance (TA) of the radio signal. In SPM studies in Estonia we use the Cell Global Identity + Timing Advance (CGI+TA) method for positioning (Figure 1). Depending on the type of base station (omnicell or sector cell) and the distance from the antenna, the positioning result will be either a sector/arc or a circle. This positioning sector with a standard width of 550 m is the main characteristic for using positioning data and calculating error. The closer the phone is to the antenna, the smaller the positioning inaccuracy, i.e. error. The proximity of an antenna depends on the density and load of the network and the location of the mobile phones.

2.2 Collecting positioning data

Collection of mobile positioning data is technically possible in the majority of mobile networks. The positioning method and data quality (accuracy) can vary a lot because of hard and software differences. The biggest problem in organizing positioning experiments is the willingness of operators to offer the infrastructure, as surveillance fear is a very sensitive matter. Operators do not want to lose their clients' trust as surveillance and tracing experiments can generate fear. Still many operators accept LBS-s and experiments as a potential source for generating additional income from the expensive infrastructure. Mobile positioning data is similar to movement data gathered by different means such as travel diaries, GPS, Moby drive, questionnaires, video tracking, etc: two or three location coordinates (x,y,z) and time (t). Therefore, SPM survey is not a revolutionary data collection method, but does have some advantages:

a. possible to save more movement points than traditional diaries, observations or questionnaires;
b. data is originally digital, free from respondents memory bias or manual digitalisation errors;
c. data is collected by a third body (network operator), independently from respondent's or researcher's intentional bias;
d. mobile phones are widespread and most people like to carry their mobile phone with them at all times and recharge the battery.

Positioning data also has disadvantages such as the tracing (transparency) fear of respondents and problems with accuracy. Another problem is the relatively high cost of experiments today because of the large quantities of data (location points) collected. The cost of a data unit is still lower than for questionnaire surveys.

Another important problem of positioning experiments is the technical readiness of operators to position large samples of respondents with a dense temporal interval. In Estonian SPM studies we have developed a special online environment for positioning queries to handle larger samples and dense queries. This environment is also used for collecting personal data (profile) from respondents and special aspects of travel behaviour. For the collection of mobile positioning data in SPM experiments we used standard queries which were determined by the following parameters:

a. phone number to be traced;
b. personal permission code: double permission (www questionnaire and sms message) of respondent to trace their movements during a certain time;
c. positioning query dates and frequency;
d. additional questions about personal profile and travel diary.

For collecting this active positioning data we used the special internet based positioning environment www.positium.ee which has been developed by Positium ICT (Positium 2006) and has a direct (secure) data exchange channel with the positioning servers of two major Estonian mobile operators (EMT and Elisa) (Figure 2). For positioning in the www.positium.ee environment double approval from the phone user is needed. For this purpose positium.com developed a special secure verification tool which combines Internet queries and sms messages. The double check allows mobile operators to be sure that data collection (positioning) can be done only with the personal permission of the phone user.

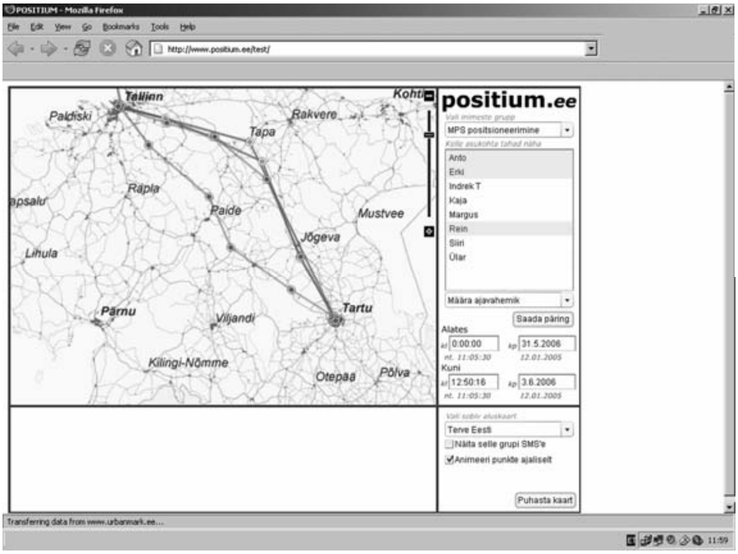


Fig. 2. Internet based positioning environment www.positium.ee used for authorization of respondents and for managing active tracing experiments. Respondents can observe their own location via Internet during studies with handled passwords.

The collected data is stored in an independent database server, which provides output to users. The respondent's personal data and codes are stored separately according to the regulations of personal data protection.

2.3 Positioning data

According to the positioning technique, the CGI+TA method normally gives the location of a mobile phone as a sector (Figure 1). The Ericsson MPS system delivers location parameters of those positioning sectors in the following format: coordinates of antennae; angle (width) of positioning sector; inner radius; outer radius (Figure 3). This information is used for many LBS applications directly; other applications need the calculation of polygon coordinates or single point coordinates as the location of a mobile phone, which makes calculation graphically clear. In SPM studies special GIS application software were used to request for locations of mobiles from the mobile network and to calculate coordinate points (x,y). These co-ordinate points were stored in a table as polygons, normally one location has 4-50 polygon vertical points. The elevation (z-coordinate) could not be determined using this positioning technology.

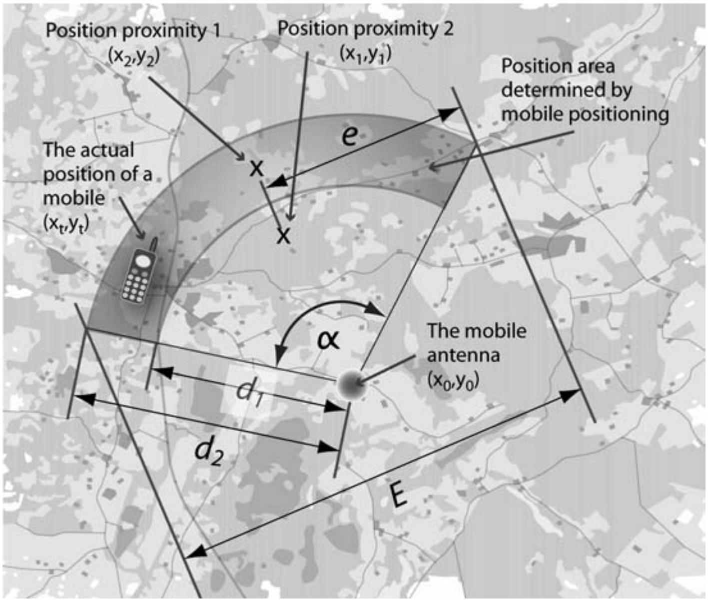


Fig. 3. Parameters of "positioning sector" gathered by CGI+TA and used for LBS applications. d1 – inner radius of positioning sector; d2 – outer radius of positioning sector; e – expected maximum location error; α – positioning sector angle, coverage limits of network cell

Positioning sector parameters are too complex for practical location calculations, for modelling a person's movement or calculation of distances. For these purposes the position must be approximated to a point location. Different LBS applications and GIS software have different methods to calculate the central point of a positioning sector which is actually the approximated (estimated) location point of phone. We used two parallel methods for calculating the central point of a positioning sector. The first method determined the mean of the X and Y co-ordinates describing the positioning sector – Position proximity 2 in Figure 4. The second method we used determines the area centroid (gravity centre) of the positioning sector using standard GIS tools – Position proximity 1 in Figure 4. These different methods give slightly different locations and are used by different end users.

3 Data

All surveys described here were conducted with customers of an EMT labelled GSM network in Estonia. The land area of Estonia is 45 000 square kilometres and approximately 70 per cent of the 1.4 million population live in urban areas. The population pattern determines the location and density of the mobile networks. The EMT network (biggest in Estonia by subscribers and radio coverage) used in the study has offered its mobile positioning services (LBS) since the year 2000. The basic technical characteristics of the positioning network are the following: a nation-wide GSM network using the 900 MHz frequency band, additionally 1800 MHz and 3G in city centres; the total area covered is 90 000 square kilometres, and the network includes approximately 900 base station sites. The network used in the present work includes an MPS (Mobile Positioning System) based on Ericsson's GSM network and using the CGI+TA method (Cell Global Identity and Timing Advance) for MPS mobile positioning (Ericsson 2003). The network of the biggest Estonian Network operator EMT used in SPM experiments is described in Figure 4.

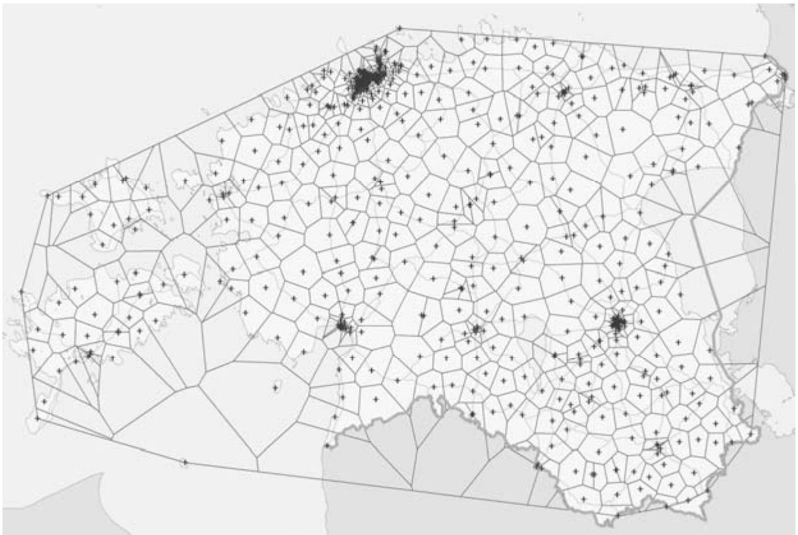


Fig. 4. Network cells determined by Thiessen polygons surrounding base stations (marked with +) of biggest Estonian operator EMT network. EMT uses Ericsson positioning hard- and software.

1. The network preciseness was calculated on the basis of the positioning characteristics of the EMT network in Estonia.
2. The estimated positioning error was determined using evaluation data for calculations based on actual positionings carried out in the EMT network, in total 180 000 measurements of 27 000 mobile phones made over a 9 month time frame during 2003. Data was collected passively: positionings were made during everyday usage of different mobile location-based services provided by the network, mainly *"Friend Finder"* and *"Find Nearest"* service. This data included a timestamp of positioning, the location area identified with the CGI+TA method (MPS) (>95 per cent probability), but not the actual position of the phone.
3. The SPM experiment was carried out by Positium Ltd and the Institute of Geography of the University of Tartu in the metropolitan area of Tallinn from February 18-22, 2004. This material was used for space-time behaviour analysis for city planning. A sample of 117 people were positioned with MPS (CGI+TA) during 5 days with 30 min intervals; all together more than 14 000 positioning points were collected. During the experiment 717 reference measurements were made with GPS.

4 Accuracy of mobile positioning in the Estonian GSM network

4.1 Network density and positioning bias

The main problem with the CGI+TA method is its relatively low accuracy of positioning. In practice, the accuracy (precision) of mobile positioning depends to a large extent on the density of the network, inter-distance of base stations/antennas. Mobile network density is normally correlated with population geography: cities and dense population areas have the most dense network infrastructure and therefore also a good accuracy of positioning. Figure 3 shows potential positioning accuracy in the EMT network in Estonia calculated from the network parameters: size of positioning sector and distance from base station.

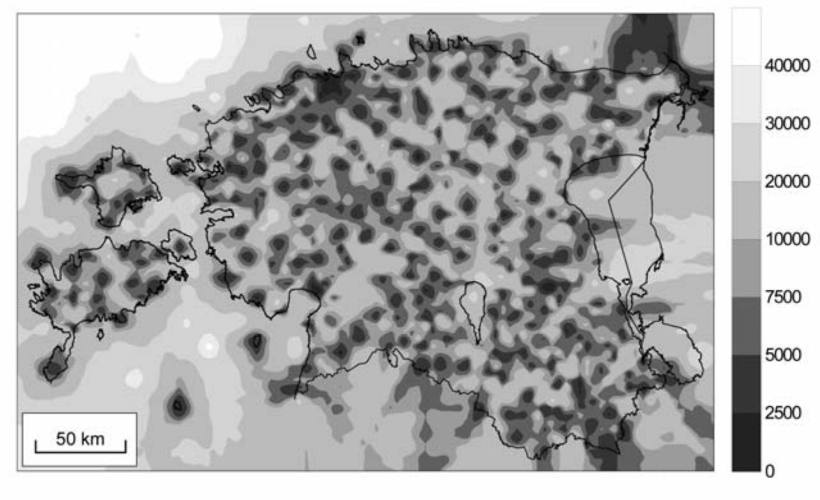


Fig. 5. Accuracy (meters in right scale) estimations of mobile positioning in EMT network in Estonia. Calculated on the basis of radio coverage parameters, positioning method CGI+TA.

Mobile positioning using CGI+TA has some technical specifics. One is the positioning bias caused by inter-cell hand-over, meaning the occasional switch of a telephone from one network antenna (i.e. cell) to another, when both cells have good coverage in the same spot. A person and phone can stay in the same location, but the network switches the mobile radio connection to the second antenna, which has a greater capacity or its signal level is slightly better. So even when the user's actual location has not changed, the calculated position of the person can change significantly. This bias is called *"hand-over noise"* in SPM experiments.

Another source of error typical for the CGI+TA method is correlated to distance from the network base station antenna. If the distance is too long, the error will grow exponentially. This case is especially problematic in less populated areas and coastline areas where networks use antennas with extended coverage. Also in cases when the closest base stations to the telephone are overcrowded, the mobile phone can be connected to another base station which may be farther away from the user. This can also create a greater positioning error. The maximum distance from a regular GSM base station to handset is 65km.

4.2 Calculated positioning bias

Additionally, we use the parameter called the calculated positioning bias calculated with base station distribution information and 180 00 actual positioning points from the EMT GSM mobile network. Figure 3 shows the calculated positioning bias as (e) and its value for each positioning can be found with the formula e2=d2 * sin (α/2). The average estimated bias (*e*) for the 180 000 positioning measurements was 2 540 metres. The minimal error in this data set was 465 m and maximum 62 km with a standard deviation σ = 3069.95. As stated earlier, the estimated error depends mainly on the distance of the phone from the base station, the density of the network and the geographical distribution of the phones. To a certain extent, the possible accuracy is also affected by features of the landscape affecting radio coverage such as relief, buildings and vegetation. For reducing side effects and increasing accuracy EMT is using special software Pin-Point Mgine (Regio 2006).

Table 1. Cumulative fractions of the theoretical positioning bias (calculated on the basis of network parameters) on the example of 180 000 positioning points carried out in Estonia.

Error e (m)	Number	Fraction	Cumulative fraction, total	Cumulative fraction in urban areas	Cumulative fraction in rural areas
<500	6774	4%	4%	8%	1%
500-1000	50508	29%	32%	61%	10%
1000-1500	36846	21%	53%	89%	25%
1500-2000	16332	9%	62%	96%	37%
2000-2500	10437	6%	68%	98%	45%
2500-3000	8689	4%	73%	99%	53%
>3000	47559	29%	100%	100%	100%
Number				77 176	99 969

As indicated by the cumulative fraction of positioning error in Table 2, the theoretical accuracy of 89 per cent of positionings studied remains below 1500 metres in urban areas while it is greater than 2000 metres for 55 per cent of positionings in rural areas.

4.3 Comparison of mobile positioning and GPS data in 2004

Positioning error can be directly determined as the difference between the CGI+TA measured location (converted to calculated location point) and the actual location of mobile phone, as measured using GPS. For this purpose, the positions of the mobile phones were identified in parallel to mobile positioning, during our field experiments with GPS device. A total of 717 measurements were made during the experiment, which while not a great number, was nevertheless enough to validate the results of the experiment. The GPS and GSM positioning results were chronologically synchronized, coordinates were transformed to a projected co-ordinate system and the average error was calculated using simple orthogonal coordinate calculations $e = sqrt\ (dx^2 - dy^2)$. Figure 6 shows an example of the differences between mobile positioning and actual (GPS) positions in urban and suburban areas.

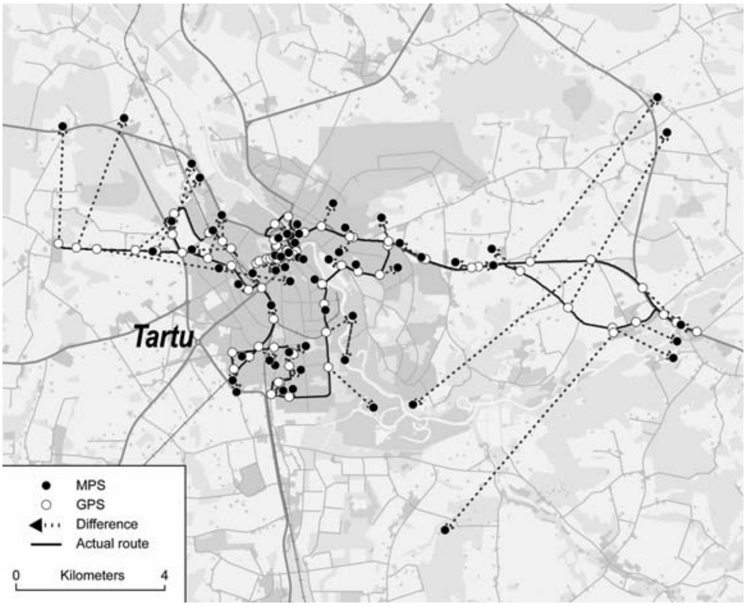


Fig. 6. Differences in positions (MPS) given by mobile positioning and GPS, in the example of positionings with 5-minute intervals in Tartu, Estonia.

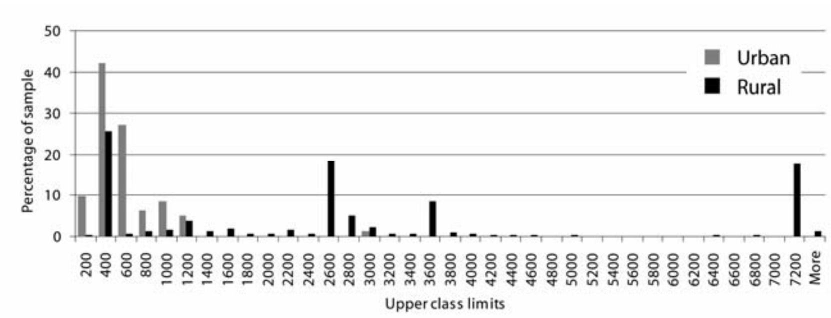


Fig. 7. Distribution of differences between MPS and GPS measured points in urban and rural areas.

The distribution of positioning error is also shown in the figure 7, compiled separately for measurements in urban and rural areas. The figure shows the distribution for error categories in steps of 200 meters. As the number of measurements differed in urban and rural areas, the values on the figure are given in percentages. The distribution of positionings (Figure 7) shows clearly that MPS is much more precise in urban than in rural areas. 52 per cent of the measurements conducted in urban areas have accuracy better than 400 metres and 99 per cent better than 1200 meters.

The distribution of actual errors in rural areas is considerably more dispersed. The majority of errors fall into four categories: 200 - 400; 2400 - 2600; 3400 - 3600 and 7000 - 7200 meters. The largest category, with 26 percent of the measurements, is 200 - 400 meters, the same as for urban areas. The accuracy of 50 per cent of positionings carried out in rural areas is better than 2600 meters.

While the positioning error in urban areas remains below 400 meters (2600 meters in rural areas) in most cases, as shown above, it is interesting to note that a considerable portion of the actual positions, as measured with GPS, are outside the mobile positioning sectors. In fact, out of these 717 positionings, the GPS point hit the mobile positioning sector in only 463 cases (55 per cent). The reason for such a low precision is the method used for finding centroids in the positioning sector. This is driven by the market demand for clear service - everyday users of the LBS services require as small a sector as possible. Many end-users need one point as a location mark. The correlation between the size of the positioning sector and the difference from the GPS-identified point (error) was studied. There is a strong correlation between the size of the sector and the positioning error: the larger the sector, the greater the positioning error. The value of Spearman's rank order correlation coefficient for the 717 measurements was 0.84 (Figure 8).

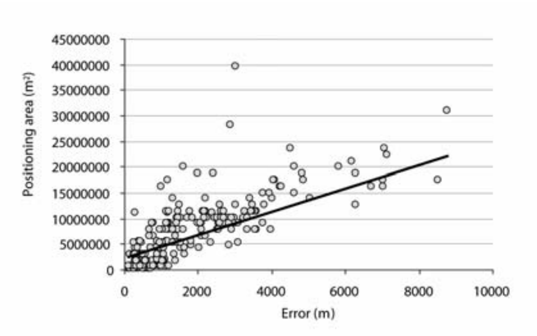


Fig. 8. Correlation between size of positioning sector (area) and the positioning error of MPS, on the basis of the 717 measurements made with GPS.

4.4 Comparison of two methods for calculation of central point of positioning sectors

For improving the accuracy of MPS based positioning calculations it is important to analyze the accuracy of the two methods we used for the calculation centre of positioning sectors i.e. the calculated position of phone. We used two different methods of calculating the point positions in the positioning sector: XY mean and the centroid. For accuracy of calculations we used the reference data of the same 717 GPS measurements.

Results show that using the mean of the XY coordinates is the more accurate method. Using the XY coordinate mean method, the positioning error, compared to the centroid method, was smaller in 70.1 per cent of cases. The difference is shown in figure 9. The actual differences in the mean values however, are not very large, reaching about 100 metres, or less than 4%.

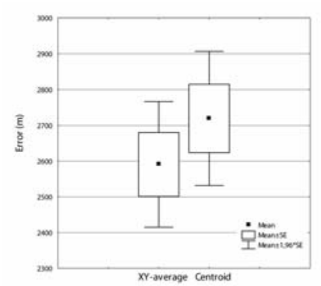


Fig. 9. Difference of the mean errors achieved with two different methods from the position given by GPS. Left bar: mean of the XY coordinates; right bar: and centroid.

5 Discussion and conclusions

The results of our experiments and analysis indicate that data from mobile positioning can be used for geographical research and human behaviour, even with the limited accuracy that is achievable today. As most of the surveys were conducted in the Tallinn and Tartu urban regions where the network is dense, the study results confirmed the achievable positioning accuracy of 500 metres. A majority of the 14 000 positioning measurements made in Tallinn remained within the accuracy range of 600 meters. This result was adequate for the analysis of commuter's daily movement and use of recreation areas. The analysis of the use of the urban space was very easy for larger city districts, suburbs and entire urban regions. Also the data analysis and accuracy was sufficient to analyze the use of the 300- to 700-metre transportation zones determined in the city. Problems were encountered when attempting to link the data to specific blocks, streets or buildings, i.e. with 100- to 200-metre precision. This is the level that architects and developers normally want to achieve for making detailed plans and houses. GIS software packages are capable of automatically assigning a street canyon along which the line between two positionings runs, but the supplementary GPS measurements made during experiments revealed that the accurate identification of a street used by our respondents is difficult in a normal urban area. Similarly, there are difficulties in linking a person's position to a specific block, street or house. For that purpose, our questionnaire (travel diary) for participants in the experiments included questions about home and business addresses, transportation routes and devices. Most people in our sample spent more than 80 per cent of the day at home or at work and with this additional information the identification of the positions by blocks and houses was simple. All studies use such questionnaires for determining the participants' personal characteristics and this could be part of the standard positioning contracts people have to fill in to give access to positioning data. Without written consent from the subject, use of mobile positioning information is prohibited by law in Estonia and in most other countries as well. A more difficult task than identifying the work and home address is the identification of the exact route between two positionings. If geographers, planners or architects want to know exactly who uses one particular street and when, it may be difficult to answer with today's accuracy of positioning. However, identification of movement on highways and major streets is much easier as they are longer and the GSM base station antennae are normally directed along the highway axis. Identification of routes is also easier in rural areas where the accuracy of positioning is lower but the number of possible routes to select from is smaller.

The use of positioning in studies of the temporal rhythm of society is very successful. This addresses especially the study of diurnal, weekly and seasonal rhythms of society or traffic. The 14 000 positioning measurements of our experiment in Tallinn indicated a high correlation with the automatic traffic counters at the main intersections in Tallinn. The correlation between movements of all SPM data and the traffic at main intersections was normal (R=0.63). The traffic flow and SPM (persons) flow were displaced by 1–2 hours, which is logical.

Researchers and end users are naturally interested in as accurate a data as possible. There are three main aspects to making positioning measurements more accurate. First, the accuracy will increase if the technical network equipment is improved. Networks can be more dense and exact. Positioning requirements can be taken into account when setting up and directing antennas. However, the greatest technical hopes are on mobile positioning based on A-GPS. A-GPS incorporates the GPS positioning to improve the accuracy and the resulting accuracy can be down to 10 meters (GALILEO 2006). A-GPS is already used in some mobile networks (primarily in the USA and Japan) and it will also be possible to use the European alternative, A-GNSS, which will be based on the new satellite navigation system Galileo, to be launched within the next few years (GALILEO 2006).

Second, more advanced mobile positioning methodology like E-OTD and E-CGI (Adams et al, 2003) may also increase the accuracy of positioning using the existing infrastructure. Better methods of positioning have existed for some time already but their deployment is slow as there are not many users of location-based services, and the more accurate methods require greater investments to technological equipment in mobile networks. At the same time, the reason for the small market is the low accuracy of the deployed positioning technologies.

Finally, positioning can be made more accurate by introducing supplementary questionnaires and interactive travel diaries in mobile phones. The questionnaire used in the survey in Tallinn in 2004 allowed identifying the position for almost 80 per cent of cases, once the home and work addresses were known. Furthermore, questionnaires may provide additional data on the respondent’s daily movement habits and routes. The use of mobile phones for positioning additionally offers a direct phone or SMS channel to the respondent, which has great potential for geographical surveys.

Acknowledgements

This paper was funded by the target funding project 0182143s02 of the Estonian Ministry of Education. The authors express their gratitude to the Estonian mobile network operators, participants in the experiment, EMT, Positium, Regio and the Planning Department of Tallinn for their co-operation.

References

Adams PM, Ashwell GWB, Baxter R. 2003. Location-based services – an overview of standards. BT Technology Journal, 21(1), 34–43.

Ahas R, Mark Ü. 2005. Location based services – new challenges for planning and public administration? Futures, 37(6): 547-561.

Ahas, R., Aasa, A., Mark, Ü., Pae, T., Kull, T. 2006. Seasonal tourism spaces in Estonia: case study with mobile positioning data. Tourism Management (Published online July 2006).

Ahonen S. Eskelinen P. 2003. Mobile terminal location for UMTS. IEEE Aerospace and electronic systems magazine, 18 (2): 23-27.

Asakura, Y. and Hato, E. 2004. Tracking survey for individual travel behaviour using mobile

communication instruments, Transportation Research C 12: 273-291.

Ericsson 2003. Mobile Positioning Guide. On-line: http://www.ericsson.com/ mobilityworld/

Laineste L. 2003. Mobiilpositsioneerimise täpsus ja rakendatavus. Institute of Geography, University of Tartu, Master Thesis, 75 p.

Mannings R, Pearson I. 2003. 'Virtual Air' – A novel way to consider and exploit LBS with AR. Journal of the Communications Network, 2(1): 29–33.

Mountain, D. & Raper, J. 2001. Modelling human spatio-temporal behaviour: a challenge for location-based services. Proc. of the 6th Internat. Conference on GeoComputation (pp 24-26). Brisbane, Australia: University of Queensland.

Ohmori N, Harata N, Nakazato M 2005. GPS Mobile phone-based activity diary survey. The 6th International Conference of Eastern Asia Society for Transportation Studies - Gearing Up for Sustainable Transportation in Eastern Asia. Bangkok 21 -24 September.

Positium 2006. Social Positioning System. www.positium.ee. Last visited 28.08.2006.

Regio 2006. ReachU Pin-Point Mgine. www.regio.ee. Last visited 28.08.2006.

Spinney JE. 2003. Mobile positioning and LBS applications. GEOGRAPHY 88(4): 256-265.

A Framework for Decision-Centred Visualisation in Civil Crisis Management

Natalia Andrienko, Gennady Andrienko

Fraunhofer Institute AIS
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
natalia.andrienko@ais.fraunhofer.de

Abstract

This paper describes a framework for intelligent visualisation to be applied in emergency situations for supporting the crisis management personnel and for informing and instructing the population. The research and development work on this topic is being done within OASIS – an Integrated Project focused on the Crisis Management part of the programme "improving risk management" of IST (IST-2003-004677, see http://www.oasis-fp6.org/). Conducted over 48 months (started in September 2004), OASIS aims at defining a generic crisis management system to support the response and rescue operations in case of large-scale disasters. The project is coordinated by EADS (France). The work on intelligent visualisation is a part of the subproject dedicated to applied research and development of innovative tools for advanced decision support in emergency situations [1].

Intelligent visualisation that supports decision making in emergency situations can be viewed as an instantiation of the concept of decision-centred visualisation (DCV), which has been introduced in mid-nineties in the context of US military contracts. Essentially, DCV means the usage of problem-oriented domain knowledge for intelligent data search, processing, analysis, and visualisation in time-critical applications [2]. An ultimate goal of DCV is to optimize the overall decision-making process

and improve the quality of decisions on the basis of intelligent visualisation and related scientific disciplines and technologies. DCV is usually applied in conflict domains such as military mission planning, counter-terror intelligence etc. DCV is used in these areas as a core component of modern control rooms. Public sources point to significant investments to the DCV development [3].

Within OASIS, we are going to apply, adapt, and extend the DCV concept to the tasks arising in civil protection and crisis management. In particular, intelligent visualisation is planned to be used not only for supporting decision makers in control rooms but also for alerting, informing, and instructing the on-site personnel, the population at risk, and the general public (through mass media).

1 Problem statement

The objective of intelligent visualisation may be formulated as "give everybody the right information at the right time and in the right way". Two parts are present here:

1. "Right information at the right time" means that a person or organisation should be able to get the information that is necessary for the adequate behaviour in the current situation or fulfilling this person's or organisation's tasks. The relevant information should be provided at the time when it is needed.
2. "In the right way" means that the information should be presented in a way promoting its rapid perception, proper understanding, and effective use.

The first part refers to the problem of the selection of the relevant information, depending on the situation and the needs, goals, and characteristics of the intended recipient. The second part refers to the problem of effective preparation, organisation, and representation of the relevant information. This, again, depends on the goals and characteristics of the intended recipient and should take into account the specific constraints of emergency situations, in particular, the time pressure and stress factor. The general requirements include:

- Reduce the information load on the recipient: not only irrelevant information should be excluded but also the relevant information should be adequately aggregated and generalised leaving out unnecessary details.

- Increase the clarity of the information presentation so that the information can be appropriately understood despite of the emotions and distractions involved in an emergency situation.
- Use such methods of representation that allow quick recognition of the meaning of the information conveyed.
- Take into account the characteristics of the medium used for receiving the information.

Intelligent visualisation supposes that both the selection of the relevant information and the subsequent processing, organisation, and representation of the selected information are automated. This may be done by applying the knowledge base technology, i.e. incorporating expert knowledge in the visualisation software.

2 Actors and their information needs

To be able to select the information relevant for a specific addressee, the software must understand the needs of the addressee. The needs depend, among others, on the *role* of the addressee in the emergency situation. The following generic roles may be considered:

- **Analyst**: a person (typically in the situation control room) that needs to understand the current situation and its development, identify problems, and find proper ways of solving the problems.
- **Decision maker**: a person who chooses a specific action course to take in the current situation on the basis of an overall view of the situation, identified problems, and possible ways of solving them. In principle, one and the same person can play the roles of both an analyst and a decision maker.
- **Planner**: a person who builds a plan for realising a chosen way of solving a problem or achieving a specific goal, with assigning tasks to performers and allocating available resources to the tasks. Again, this may be the same person as the analyst and/or decision maker.
- **Performer**: a person, group, or organisation fulfilling a particular task or sequence of tasks. A performer may need to make various tactical decisions depending on the specifics of the situation and its changes.
- **Sufferer**: a person, group, or organisation that is exposed or may be exposed to some of the danger factors of the emergency situation.
- **Observer**: a person or organisation that is not directly involved in the emergency situation but is interested in receiving information about it.

This includes, in particular, the mass media, which distribute information about the situation among the general population.

The knowledge base of the intelligent system should specify the typical information needs for each of the roles. As an example, let us consider the needs of an analyst. An analyst deals with two sorts of information:

- **Situational data** characterising the current emergency situation and its progress over time.
- **Reference data** (long-term) characterising the area where the situation occurs, the objects and population in this area, the road and service infrastructure, etc.

An intelligent system can help the analyst by automatic retrieval of relevant reference data on the basis of situational data. For example, the system can automatically find the objects located in the danger zone and requiring particular attention as potential sources of danger (e.g. a petrol station close to a fire or a chemical factory in an area that may be flooded) or valuables that must be protected (e.g. a historical building, library, or art gallery). The system can also retrieve data about the population in the danger zone, estimate the total number and the number of persons of the most vulnerable categories such as children, elderly, and disabled. It can also find the locations of people that may need special care: hospitals, schools, kindergartens, homes for elderly, etc., taking into account the current day and time. Thus, the system can look for the data concerning schools and kindergartens only in a case when the emergency situation occurs on a weekday during the working hours. The system can also find buildings or places outside the danger zone that can be used for sheltering people or animals, and so on.

Unlike an analyst, a decision maker does not need detailed information about the situation and every item requiring attention. For example, for making the decision whether to evacuate the population from the danger zone and, if yes, in what way (centralised or allowing people to evacuate on their own using the private cars), the decision maker needs to know the number of population at risk, the risk probability, the potential consequences of no action, and the time and resources required for each variant of evacuation. This information can be prepared by the analyst with the help of the intelligent system, and the system can present in an adequate form.

A planner does not need all information concerning the current situation but only the information relevant to the particular problem the planner needs to solve at the moment. For example, the planner's current goal may be to build a plan for the evacuation of the patients of a particular hospital.

For this purpose, the planner needs the information about the location of this hospital, the number of patients and their specific needs (e.g. special equipment may be necessary for the transportation of some patients), the hospitals outside the danger zone, their profiles, and the availability of beds in them, the available roads, transportation means, etc. This information can be automatically retrieved and appropriately presented to the planner.

A performer needs the information relevant to the task to be performed. This includes the description of the task, the place, the time frame, and the resources and infrastructure to use. This information is prepared by the planner while the system can take care of organising and appropriately presenting it. Additionally, the system can automatically supply the performer with the information as to how to get to the place, how to protect the personnel, what sources to use for renewing consumed resources (e.g. water or fuel), and what roads are destroyed or blocked, etc.

A sufferer may need information on how to get out from the danger zone, how to protect himself and/or his property, what kind of help is available and how to get it. Actually, a sufferer may simultaneously be a performer when he/she is supposed to follow certain specific instructions, for example, concerning the evacuation. In this case, the sufferer needs the task-relevant information: what to do, where, when, and how. The information can be automatically retrieved and prepared individually for each sufferer/performer on the basis of the general evacuation plan. The system can take into account the location of the sufferer and his specific (dis)abilities and needs if this information is available or can be inquired from the sufferer. For example, the system can send an alarming and instructing message to sufferer’s mobile device and ask the sufferer to specify what kind of assistance is needed, if any.

An observer needs general information concerning the character of the event (e.g. fire or flooding), its location, history and development forecast, number of casualties, and the measures undertaken for managing the situation. The information can be automatically retrieved, summarised, and suitably represented. The system can take into account the location of the observer in relation to the emergency site. Thus, if this is the same city or town, the observer is interested in the exact location of the event, which can be shown on a map of the town or city district. For an observer from another town in the same country, it is sufficient to indicate the location of the incident on a country map while an observer from another country should be informed where the country in which the incident occurred is situated.

For satisfying these different information needs, various output media can be used. Thus, for situation analysis, decision making, and planning, the emergency management personnel may use a big screen in a control room or a standard screen of a desktop or laptop computer. Performers on site may be informed and instructed by means of hand-held or head-mounted devices but also on paper (e.g. by fax). People in the danger zone can be alarmed, warned, and instructed through their mobile phones and electronic information boards while information kiosks may provide additional information when appropriate. General public (observers) is usually informed by means of TV and newspapers. Each type of media imposes its specific constraints on how information can be presented and further dealt with. The intelligent visualisation support system must take these constraints into proper account.

3 Knowledge

Let us analyse what sorts of knowledge are needed for an intelligent software system providing visualisation support to people and agencies involved in emergency situations.

In order to provide the intelligent services outlined in the previous section, the system must understand the meaning (semantics) of the situational and reference data. For this purpose, the system must be aware of various general notions relevant to the domain of emergency management and the relations between those notions. Such a system of interrelated notions is usually called domain ontology. The general ontology of the domain of emergency management is not linked to a particular emergency situation or geographical area. However, the ontology may be used for indexing various situation- and area-specific data, which allows the system to “understand” the meaning of these data.

To enable the understanding the situational and reference data pertinent to emergency management, the domain ontology must contain several categories of knowledge. One of them describes the **danger factors** that may be involved in emergency situations: fire, explosions, water, wind, contamination of air, water, or soil, disease, etc. For each possible danger factor, the following general knowledge is needed:

- Existential behaviour: does the danger factor effect momentarily (e.g. an explosion or earthquake) or over an extended time period (e.g. a fire or contamination)? Does it ceases by itself or has to be stopped? Can the danger re-occur?
- Spatial characteristics: is the danger factor localised in space or extended over an area or dispersed (i.e. occurs in many disjoint locations)?

- Changes of the spatial characteristics over time: static (no changes), moving (change of the position), spreading or shrinking (change of the extent), dispersing or concentrating (change of the distribution).
- Attributes characterising the danger factor, e.g. the speed of a wind, the level of water, or the concentration of a dangerous substance.
- What object categories can be affected: people, buildings, roads, service infrastructure, environment, … For each potentially affected item, the following information needs to be known:
  - Possible results of being affected: destroying, damage, loss of utility, transformation into a danger factor, …
  - Dependance of the effect on the spatial and temporal distance from the danger and on its attributes.
  - Possible measures to prevent or mitigate the effects of the danger, e.g. removal (evacuation) from the danger zone or protection.
  - What to do if the effect already took place, e.g. repair, remove, restrict access, neutralise, …

Another category of knowledge describes the generic tasks that are often involved in emergency management, such as evacuation of people, animals, and valuable objects from the danger zone. The system needs to understand what sorts of places or buildings can be used for sheltering various categories of items, for example, that patients from a hospital must be moved to other hospitals with appropriate equipment, depending on the state of a patient. The system must also know what categories of people need special means of transportation and/or assistance for embarking and disembarking.

To support decision making and planning, the ontology must contain knowledge about the types of resources and infrastructure that may be needed for managing emergency situations. This includes people, teams, and organisations (e.g. a fire brigade or a bus company), transportation means, roads, sources of power, fuel, and water, and so on.

As we already indicated, the system should also have knowledge about the possible actors and their typical information needs. The actors may be defined in terms of their roles in an emergency situation.

Besides the various categories of knowledge specific to the domain of emergency management, the intelligent system needs to know how to manipulate, organise, and present various types of data. It must be able to build different types of displays (maps, graphs, diagrams, charts, tables, etc.) and know what kind of data each display is suitable for. The system should also be able to combine several displays in a single presentation by designing an appropriate arrangement of the components and establishing links between them. It must know the properties and rules of the use of different visual variables [4]: position within the display area, size, colour, texture, etc. The system must be able to build dynamic (animated) presentations and to supply the displays being generated with various interactive facilities such as zoom, pan, rotate, focus, query, mark, filter, or follow a hyperlink.

These sorts of knowledge and abilities are, in principle, domain-independent. They can be represented and used separately from the domain-specific knowledge concerning various dangers and ways of handling them. This approach allows the visualisation knowledge to be reused for other applications. Therefore, our plan is to separate the emergency management-specific knowledge from the knowledge on visualisation design, which is specified as domain-independent. Accordingly, domain-specific reasoning used for the selection of appropriate information is separated from domain-independent reasoning in the process of designing the presentation of the selected information.

The entire intelligent system can be viewed as consisting of two cooperating expert subsystems, which may be called “emergency management expert” and “visualisation expert”. First, the emergency management expert selects the necessary information depending on the needs of the intended recipient determined by recipient’s role and the current status of the situation. Then the visualisation expert finds appropriate methods for transformation and presentation of the selected information.

4 Meta-information

For appropriate processing of the selected information and designing its presentation, the visualisation expert must take into account various characteristics of this information and how it is supposed to be used as well as the constraints imposed by the target output medium and some characteristics of the intended recipient. Let us use the term *meta-information* to refer to the factors influencing the ways of information processing and presentation. More specifically, meta-information includes the following items:

- Type of the entities the information refers to: movable or unmovable objects, places, processes, actions, or relations.
- Structure of the data and types of the components they consist of: spatial, temporal, numeric, ordinal, or categorical.

- Quality of the information: does the information result from actual measurements or observations or this is a forecast or estimation? What is its degree of accuracy or certainty?
- The goal of providing the information to the addressee:
  - **alert**, attract attention to something unexpected like an impending threat;
  - **inform**: what, where, when, how. The information may be expected by the addressee, e.g. come in response to a query;
  - **suggest**, e.g. some action to take or additional aspect to consider;
  - **enable**: analysis, reasoning, decision making, or action planning;
  - **instruct**: what, where, when, how to do or to avoid;
  - **explain** or justify, e.g. a decision made.
- Degree of relevance to the goal: information of primary importance, supporting information (e.g. orientation clues), irrelevant information.
- Degree of novelty to the addressee: completely unexpected, partly expected (e.g. an analyst may expect that some buildings could collapse in result of a earthquake but not know what buildings and where), well known and expected.
- Criticality, i.e. whether an information item requires immediate attention of the addressee.
- The expected level of addressee's knowledge concerning the topic of the information and the geographical area the information refers to.
- Characteristics of the output medium to be used for presenting the information: size, resolution, available colours, possibilities for dynamic output and user interaction, memory capacity, individual or public use.

According to our approach, the emergency management expert supplies the required meta-information to the visualisation expert together with the information to be presented. The meta-information concerning the character, structure, and properties of the information comes from the indexing of the information items in terms of the domain ontology. However, the emergency management expert must specify this meta-information in a domain-independent manner so that the visualisation expert could use it without having any domain knowledge. The meta-information concerning the goal for which the information is to be used comes from the user of the intelligent system. In some cases, the user is the same person as the intended recipient. Thus, an analyst or planner can directly interact with the system in the course of analysis or planning. In this case, the system presents the relevant information directly to the user. In other cases, the user employs the system for preparing and presenting information to other actors. Thus, an analyst may prepare a summary of the situation and possible ways of handling it for communicating to a decision maker. A performer who received the task to alert and instruct the sufferers may use the system for preparing an appropriate presentation or even let the system construct a personalised message for every individual or family in the danger zone.

So, the user somehow lets the system know for whom and what purpose the information is needed. Basically, the system needs to be aware of the recipient's role in the current situation and his/her geographical location. Sometimes, this sort of meta-information is clear from the context of the system use and does not need to be explicitly specified. On the basis of the meta-information concerning the recipient and the stated or implied goal of information use, the emergency management expert can estimate the degree of relevance, novelty, and criticality of each information item for the recipient. For this purpose, it uses the domain knowledge concerning the actors (roles) and their typical information needs. The same knowledge allows the expert to estimate the probable level of addressee's thematic and geographic knowledge. Thus, an analyst may be qualified as an expert in emergency management issues but the level of knowledge concerning the area of the incident may be low. In opposite, the local population to be alerted may know the area quite well but be unaware of the character of the particular threat and the possible consequences.

Hence, by reasoning with the use of the domain-specific knowledge, the emergency management expert not only finds and retrieves the relevant information but also supplies it with meta-information that allows the visualisation expert to interpret, process, and present this information adequately. The meta-information is specified in a domain-independent manner so that no domain knowledge is needed for utilising it. Additionally to this, the visualisation expert needs to know the characteristics of the output medium. This information either comes from the context of the system use (when the information is intended directly for the system user) or the user specifies it explicitly (when the information is to be communicated to someone else).

The organisation and functioning of the intelligent visualisation system can be schematically represented as is shown in Figure 1.

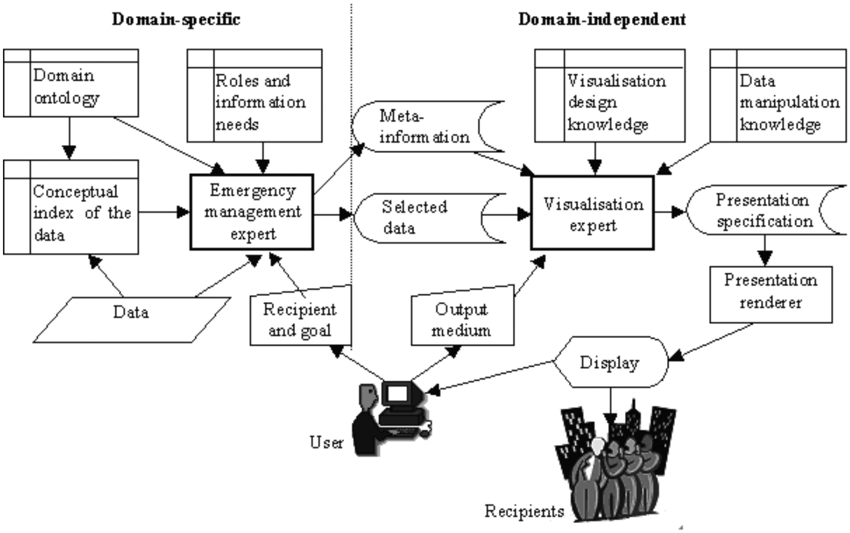


Fig. 1. A schematic representation of the structure and functioning of the intelligent visualisation system for emergency management.

5 Knowledge sources

As may be seen from the scheme, the intelligent system will use several categories of knowledge:

- Emergency management domain ontology;
- Knowledge about the roles and information needs of people and organisations that may be involved in an emergency situation;
- Knowledge on visualisation design, including techniques of interactive display manipulation and mechanisms of coordination of multiple displays;
- Knowledge concerning techniques of data manipulation, e.g. aggregation, filtering, smoothing, interpolation, statistical processing, etc.

The first two categories of knowledge are specific to the domain of emergency management. This knowledge is currently being constructed mainly by analysing available reports describing real incidents and generalising from these cases. Some comprehensive analytical reports can be found in the literature [5,6]. Much information concerning the management of real disasters can be found in the news reports from various news agencies available, in particular, in the Web. Thus, using the

Web, we have compiled a rather detailed description of the course and management of the flood in Czech and Germany in August 2002.

The second two categories of knowledge come from the extensive literature on information visualisation, geographic visualisation, data analysis, and graphics design. We have recently summarised the current state-of-the-art in these areas with the focus on techniques and tools supporting data exploration and analysis [7]. However, the intelligent system for decision-centred visualisation will also need the knowledge concerning effective information presentation according to the intended communication goals. The relevant literature is rather abundant and includes both theoretical studies (e.g. [8]) and descriptions of practical approaches to automated presentation design (e.g. [9]). Our orientation to various types of output media necessitates the use of knowledge concerning the possible ways of presenting information on these media. In particular, we can use some recent research results concerning information visualisation on mobile devices [10,11].

6 Examples

The pilot prototype of the visualisation system is now under development. At the moment of writing this paper, it is capable of retrieving relevant information items depending on the type of event (e.g. fire or flood) and the task (e.g. detect potential sources of risk or find people needing help). The presentation design capabilities are limited to building maps where impact zones of hazardous events are shown as semi-transparently coloured areas and various types of objects are represented by symbols (icons), which may vary in size depending on the level of relevance or criticality. Some examples may be seen in figures 2-5.

7 Conclusion

The realisation of the proposed framework is a very ambitious and complex task. It is hardly possible to fulfil it completely in the course of the OASIS project. Therefore, we aim at a partial realisation that would be sufficient for the demonstration of the feasibility of the approach and the potential benefits for the emergency management personnel and for the population. We are starting with one example scenario of managing a specific emergency situation and developing a prototype oriented to this scenario. On this stage, we shall restrict the scope of our work to only some of the roles and some types of the output media. After testing and refining the prototype, we shall extend it to further types of emergency situations and the remaining roles and media types.

References

[1] N. Andrienko and G. Andrienko, A concept of an intelligent decision support for crisis management in the OASIS project, In P. van Oosterom. S. Zlatanova and E.M.Fendel (Eds.) *Geo-information for Disaster Management*, Berlin: Springer-Verlag, 2005, pp.669-682

[2] J. Kohlhammer and D. Zeltzer, Towards a Visualization Architecture for Time-Critical Applications, In *Proceedings IUI'04*, 2004, Madeira, Funchal, Portugal, ACM Press, pp.271-273

[3] M. Guzzo, "MAYA Viz lands $20 million contract from U.S. Army", *Pittsburgh Business Times*, 23.7.2004, available at http://pittsburgh.bizjournals.com/ pittburgh/stories/2004/07/26/story7.html

[4] J. Bertin, *Semiology of Graphics. Diagrams, Networks, Maps*, The University of Wisconsin Press, Madison 1983, translated from Bertin, J.: *Sémiologie graphique*, Gauthier-Villars, Paris 1967

[5] U. Rosenthal, P. Hart, and M. T. Charles (eds.) *Coping with crises: the management of disasters, riots, and terrorism*, Charles C Thomas, Springfield, Illinois, USA, 1989

[6] U. Rosenthal and B. Pijnenburg (eds.) *Crisis Management and Decision Making. Simulation Oriented Scenarios*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991

[7] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach*. Berlin, Springer-Verlag, 2005 (in press)

[8] E.R. Tufte, *Visual Explanations. Images and Quantities, Evidence and Narrative*. Graphics Press, Cheshire, Connecticut, 1997

[9] S.M. Casner, A Task-analytic Approach to the Automated Design of Graphic Presentations. *ACM Transactions on Graphics*, Vol. 10, 1991, pp.111-151

[10] G. Gartner and S. Uhlirz, Cartographic Concepts for Realizing a location-based UMTS Service: Vienna City Guide Lol@, *Proc. 20th International Cartographic Conference Beijing*, August 2001, Beijing, China; vol.5, pp. 3229-3239

[11] T. Reichenbacher, The importance of being relevant, In *CD-ROM proceedings of ICC 2005 – International Cartographic Conference*, A Coruna, Spain, July 9-16, 2005

Examples

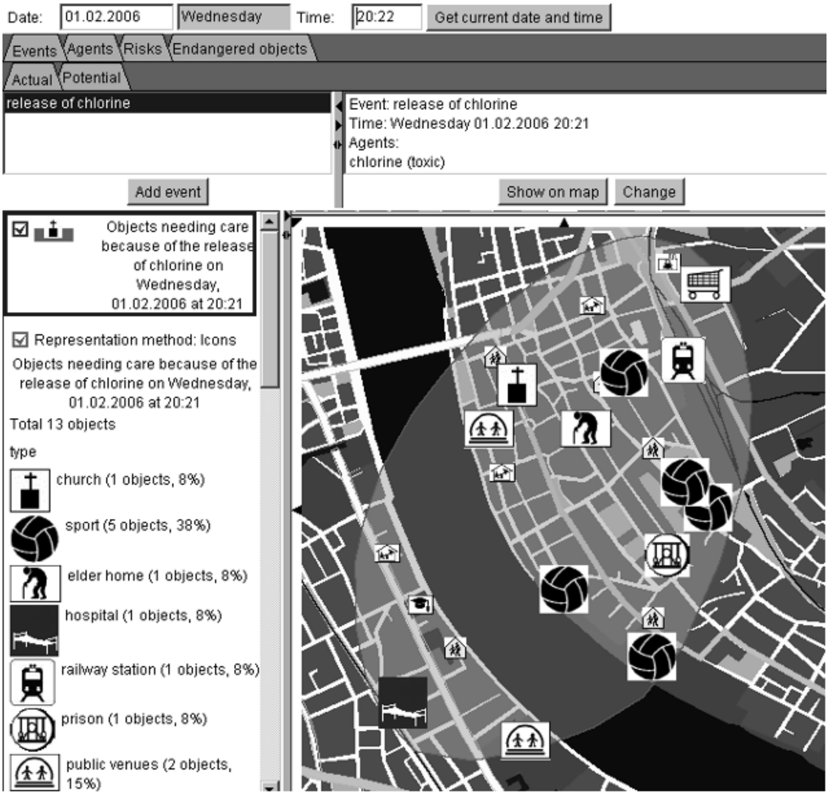


Fig. 2. The map display shows the impact zone of an event of chemical release and the endangered objects, specifically, buildings containing people that may need help. The sizes of the symbols correspond to the degree of relevance (in this case, the probability of people being inside the buildings at the specified time of the day).

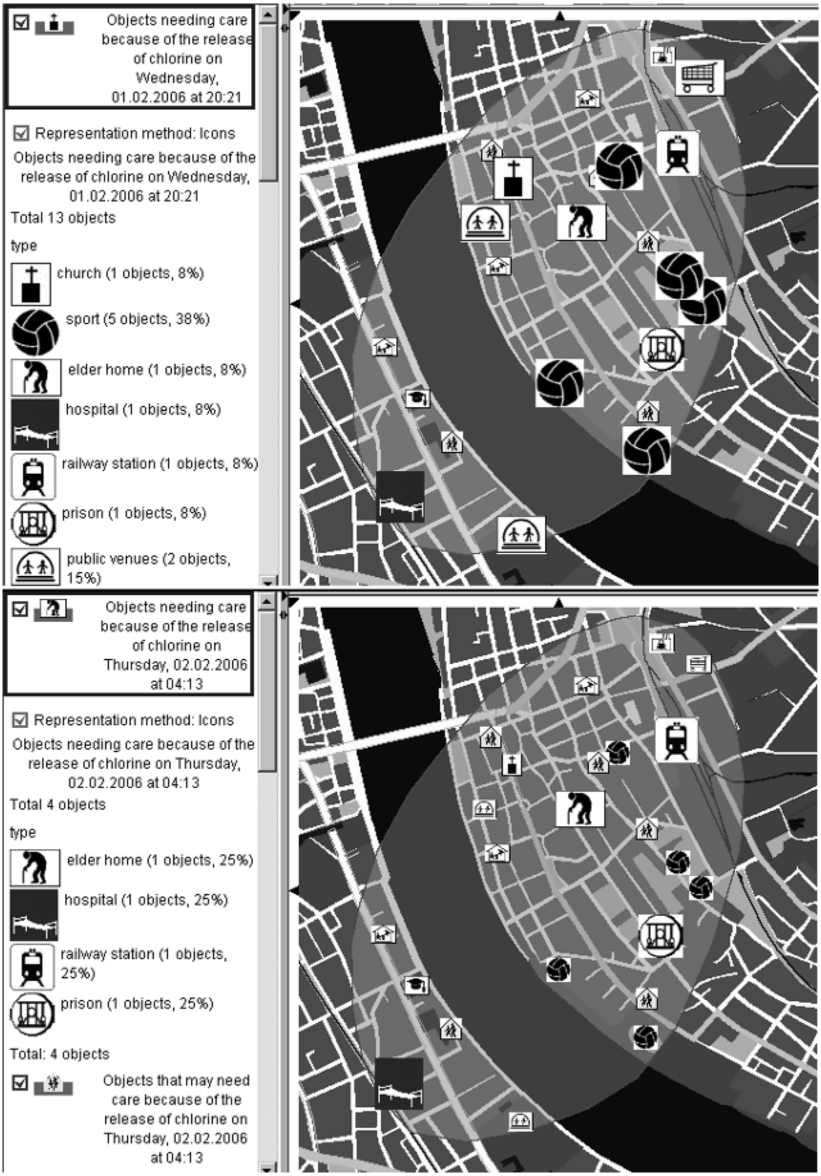


Fig. 3. Depending on the situation specifics, the same information may have different relevance. As a result, it will be given different visual prominence. This figure demonstrates how the time of event occurrence influences the appearance of the display of buildings with people needing help. The upper screenshot corresponds to a situation in the evening at 20:21 and the lower image to a situation in the morning at 4:13.

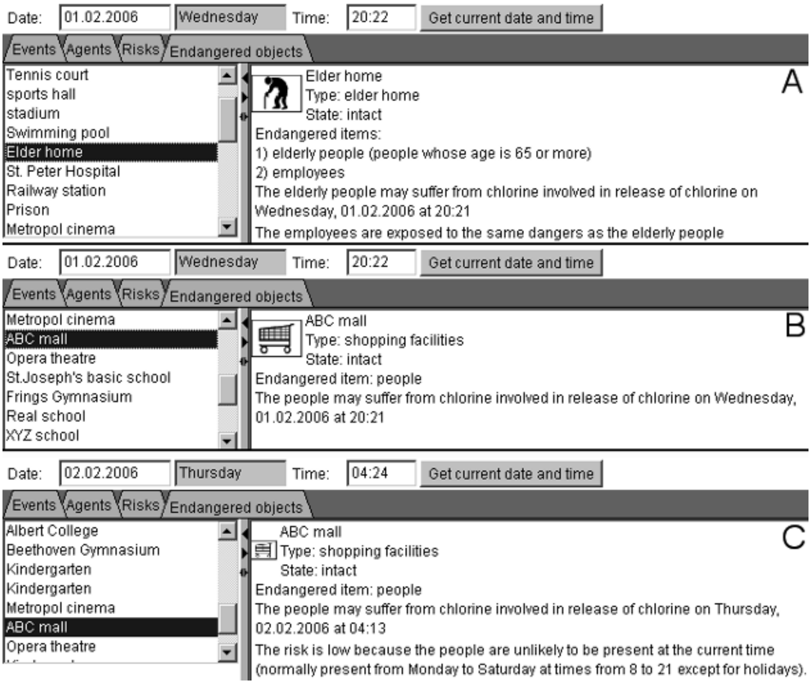


Fig. 4. The system supplies the user with explanations why each information item has been selected as relevant to the current situation. The upper two explanations (A and B) refer to two types of endangered objects, an elder home and a shopping centre, involved in the situation of chemical release at 20:21 (see Figure 2). The lower screenshot (C) corresponds to an analogous event that occurred at 4:13 AM. It demonstrates how the explanation about the shopping centre changes depending on the degree of relevance to the situation.

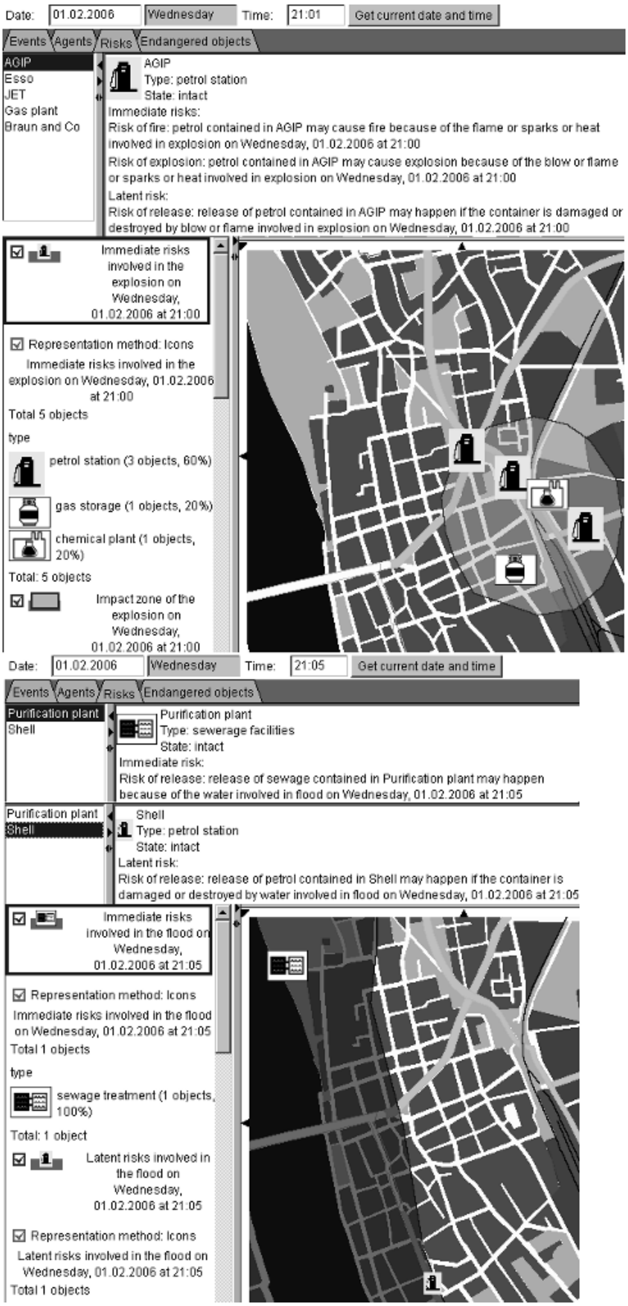


Fig. 5. Two screenshots from a map display represent the potential sources of secondary risks in the situations of explosion (top) and flood (bottom).

Providing an Information Infrastructure for Map Based LBS - The approach of the City of Vienna

Erich Wilmersdorf

City of Vienna, Hutweidengasse 7/1/6, A1190 Wien
erich.wilmersdorf@gmx.at

Abstract

Expectations of a dynamic increase in demand for mass-market LBS applications have been too optimistic. Difficulties were underestimated, because LBS are not only a technological challenge but also a complex organisational problem to solve. The major reason for expectations not being met is the lack of an adequate information infrastructure, generally about the voluminous, but reliable geo-data warehouses. Additionally, powerful analyses needs to consider both attributes of time and of space, which must be done in real-time. Finally, a cartographically-acceptable user interface is also a prerequisite for LBS, but one not so commonly implemented.

Municipalities can play an important role in this field. Their information and communication network, which is backed by the functionality of Geographic Information Systems (GIS) for day-to-day geo-related management may contribute to the design of a better IT-infrastructure for LBS as part of how they use GIS. This chapter describes the components of such an infrastructure that would be suitable for LBS applications, one that would be able to promote public and private LBS usage by offering dynamic content and standardised Web-interfaces for geo-data extraction (WMS, WFS). Finally, the approach used by the Municipality of Vienna to link GIS with LBS is outlined.

1 Introduction

Densely populated regions are marked by a high density of objects accompanied by service providing public institutions and private enterprises. Both the urban “landscape” and the services are subject to dynamic changes. Therefore the requirements for LBS are especially challenging: Information detailed not only in respect of space but also of timely aspects, is in the focus. The utmost in topicality of data with geographic reference has to be available on call, to get reliable information about the kind of service asked for, either provided by private companies or by public authorities. Urban administrations play two roles in managing space: as a producer of changes in the urban region itself (e.g. new streets) and as a controlling unit, supervising the evolution of the urban area according to the legal framework such as protection zones, building and zoning regulations. So the administrative body has the possibility to play an important role in building up a sound infrastructure based on geo-related information essential for its own tasks but also for external services as by-product, e.g. LBS.

The urban administration of Vienna has been collecting digital geo-data for three decades in its own interest: They represent a geographic knowledge data base, which facilitates monitoring and decision making in internal processes related to local agendas. On the other hand, map based Web services have been offering information to the public for the last ten years already. Therefore the administration of the City of Vienna can be seen as a service providing enterprise for citizens and economy, but also in a global respect for everybody, who is interested in information about Vienna. Former experiences of its map-based Web Services have shown the requirements necessary for LBS.

2 General requirements for a LBS infrastructure

The stand-by readiness of LBS is a tough challenge. Immediately the application must respond to a triple question: What is where and when? “When” usually means in a near and narrow time slot (e.g. travel time to a certain destination). In a processing chain starting with the request the appropriate answer to the question of the customer has to be generated instantaneously. Thus a comprehensive infrastructure is necessary for a successful LBS in operation: A powerful telecommunication network, geo-referenced data, ubiquitous computing with GIS functionality, controlled by the user over a dialogue oriented interface supported by real-time cartography (Wilmersdorf, 2003).

These prerequisites show the complexity of LBS applications and they altogether are the reason, why LBS did not fulfill the optimistic forecasts of commercial success in the past. Critical factors, lacks in the infrastructure, the costs of its implementation have been underestimated. The paper focuses on the essential part of information infrastructure, which can be made available by urban administrations.

2.1 Data requirements of LBS

LBS make high demands on the availability of data, in order to enable reliable answers to these typical threefold question: What is when and where? The need for digital data especially in urban regions is voluminous. In the focus of LBS are data in the context of location referencing:

- Geodetic position
- Geometry of objects
- Topological connexions (the spatial context for navigation e.g. network of public and private transport and finally the access for pedestrians too)

But also attributes of single objects and events are of importance, which are asked for:

- detailed classification of services,
- scope of products

Timely aspects (Bartelme, 2000) not only of spatial data but also of attribute data are of great importance due to the ubiquitous readiness of LBS. Services must have information being available dependent on e.g.

- Calendar with national holidays
- Business hours, hours of duty and especially the current status (including vacations, short-term cancellations etc.),
- time tables of public transport

2.2 GIS functionality

Geo-related analysis is the basis for spatial decision making. In the case of LBS it is necessary to adopt effective retrieval functions to select the needed raw data. After merging data of spatial and temporal relevance the concise data are handed over into the next processing stages, which are necessary for generating value-added information asked for (e.g. the nearest POI).

2.3 TeleCartography

The transmission of spatial information by means of cartography is marked as a special type of visualization, which is based on modelling focused on the perceptivity of the recipient. Thus the term tele-cartography is used in this paper for real time-cartography with special focus on the portable devices (PDA, mobile phone). That means: Ubiquitous mapping faced with incisive restrictions concerning size and resolution of the display. Maps are expected to be delivered on the fly to visualize the spatial context of objects which suits the requirements of the customer. The automated production of the cartographic output has to be provided for two tasks: The map enables navigation and geographic orientation serving as spatial reference for the thematic layer which is inserted on top of it. The design of the cartographic output has to take into consideration the additional role of the map: The spatial interface for the dialogue of the user between attribute databases and real world. The second task is the presentation of the attributes of the requested object either by thematic symbols or other media .

3 The contribution of the urban information system for LBS

Urban administrative bodies require many geographic databases to manage local and regional developments. The City of Vienna has installed geo-related processing facilities in its computer-network as an adequate way of exploiting these databases for decision making on the working place (Wilmersdorf, 2003). This infrastructure for providing information individually can be exploited for LBS as well.

3.1 Fundamental data bases for LBS

Several data bases are of importance in order to cope with fundamental requirements of LBS.

3.1.1 Meta-database

A meta-database represents a catalogue of the geo-data warehouse. It covers the specifications of the data especially with regard to space. According to those characteristics it is possible to decide if the data are suitable and also relevant for a new Location Based Service envisaged. The metadata facilitates the content management and describes characteristics of the encoded data and how they can be accessed and incorporated into an application:

- name of the data set
- reference system (coordinate system)
- object class
- object`s identification code
- attribute list
- data quality
  - accuracy of the position
  - topicality
- maintenance information (e.g. update cycle)
- exchange format(s)
- etc.

To avoid exceptional solutions the meta-database of Vienna is implemented according to international standards (ISO 19 115 specifications).

3.1.2 Geographic Identifier (gazetteer)

Localisation awareness is of utmost importance in the field of LBS. There are many kinds of positional definitions applied in every day life. They also have to be accepted as input when entering into a LBS. Apart from the self positioning by GPS or apart from the provision of the coordinates of the telecommunication cell a large scope of user defined popular terms and variants of location names must be supported. Starting point or destination may be designated by an address in a general sense: Besides the most frequently used indication names of street and house number many other popular definitions of positions such as object names (e.g. “Schönbrunn” as name of a landmark, stops of public transport) are stored geo-coded. This set of local addresses and object identifiers facilitates localisation and orientation for mobile users.

3.1.3 Geographic Reference System

The Reference System of Vienna (RBW) represents a generalized topological model of the urban structure and a digital road network, furnished with identification keys for postal addresses, blocs and street network. The postal address is geo-coded and linked twice: To the bloc and to the adjacent street segment representing its link into the street network.

These elements offer LBS logistic information. The RBW is linked with the gazetteer data base too. This forms a solid basis for examining the networks to get a suitable route from the starting point to the target point. In case of a search for an object with special attributes the nearest object of the user`s interest is recommended by taking into consideration the available network instead of the parameter “distance” only. Internally the RBW is used for traffic management (e.g. for the coordination of construction sites) of Vienna and so temporal obstacles for car drivers are available. Such reference systems can be exploited for routing even for different modes of transport (e.g. on foot, by bicycle, by car, by bus or tram,..). Street segments are furnished with attributes (one way, pedestrian zone,..), so that route guidance can be adapted to the mode. The route can be presented either decoded by an oral description (as used in car navigation systems) or by cartographic visualisation. Both useful ways for LBS.

3.1.4 Topographic data base

The Municipality of Vienna decided to build up a digital model of high resolution in the nineteen eighties. The complete urban area was surveyed by techniques of tacheometry and photogrammetry recording directly in digital mode. The data base represents a detailed description of the geography and geometry of objects. It is an outstanding coverage because of the high resolution of the model describing the complete “urban landscape” digitally by 3D measurements. The composing of 3D objects (buildings) in the data base is now under way.

Viewed by LBS it is a powerful data base: Due to its high resolution this database offers a detailed geo-coded object inventory including even small objects (e.g. kiosks, newsstands, the “street furniture” such as hydrants, pylons, telephone booths, etc.). So it can be seen as a collection of points or areas of interest comprising many object classes. The second function is important for LBS as well. The neighbourhood can be visualised in detail for local orientation by a site map derived from this database by cartographic modelling.

3.1.5 Cartographic data bases

Flexible output, especially adaptable cartographic output is essential for LBS. Urban administrations generally make use of a city map for geo-related work. The City of Vienna decided at the beginning of digital geo-information processing, to meet the demand for cartographic output primarily by automated processing on call. So in principle cartographic visualisation is marked by real time processes (see 4.2: Real Time Cartography). There are only generalized maps precompiled in a permanent storage maintained periodically.

For LBS two cartographic databases are of interest. They are used as background maps for the purpose of orientation:

City map

The city map is derived automatically by compiling data of the Reference System of Vienna (RBW), of the land use data base and/or of the topographic data base. It is continuously updated and presented by automated cartographic modelling for different scales. The cartographic visualisation is defined by a set of rules. The map processing is finished in the display stage by a vector/raster-conversion. The scope of cartographic output covers the range of 1:200 until approx. 1:20.000.

Orthophotomap

For many purposes a photographic “snapshot” of the urban landscape is advantageous if background information in a photo-realistic way combined with cartographic text is needed. Such a detailed geographic visualisation of the surface never can be provided so quickly and substantially by a map derived out of the topographic database, representing a more or less analytically generalized model of the urban landscape. But for mobile applications resolution and size of the display restrict the use of orthophotomaps.

3.2 The digital Geo-data Warehouse

The management of cities needs comprehensive knowledge about their infrastructure and the “landscape”. The City of Vienna is already equipped in addition to the fundamental data bases (see 3.1) with voluminous databases of special fields focused on the public infrastructure (buildings, road network, bridges, traffic objects, etc.). The large amount of data are stored about objects in the urban area representing a digital geo-coded object inventory of the urban micro-cosmos. This data storage is used in the daily tasks of the administration. For LBS purposes this object inventories can be exploited as well. Some examples are highlighted in the following section.

Data about many other useful POIs are available, which are installed in public interest: Libraries, play grounds, public toilets, parking zones (with/without costs) are also useful additional themes for LBS.

Data about offices of physicians, pharmacies, hospitals with detailed information about their scope of services and times of duty are available for emergency cases. Wired Internet services of the Municipality offer information by real time-cartography already. For LBS these services could be adapted.

Additionally to a broad variety of attribute data digital images of objects of cultural importance (photos, detailed site drawings) are available for tourists in a geo-coded way.

In the focus of LBS are the navigation services for mobile users: Selecting the route according to the means of transport, also considering changing possibilities between different transport modes i.e. on the basis the networks of public transport, streets, bicycle routes and especially the nodes among the different networks for changing the mode of proceeding. The public transport has to be present with its timetables for street car, bus, subway and railway. This example shows the complexity of services being essential for flexible navigation to a certain destination.

The geographic modelling of the street area as space for moving on for everybody is faced with many constraints. The Municipality of Vienna has built up a detailed inventory model for the purpose of maintenance of its streets, but meanwhile this model is used also for a quality analysis e.g. for pedestrians. So the available space of the pavement can be analysed automatically (obstacles for baby carriages such as staircases, street furniture e.g. hydrants, pylons; bottlenecks of width) or the quality of the surface (e.g. avoid cobble stone areas of the pavement, when recommending a route for handicapped people in a wheel chair).

Due to the important role of street management also short term objects and events are incorporated into the Street Data Warehouse. This database is fed by geographic announcements of construction projects in advance for the purpose of coordinating e.g. the construction sites in the road network as obstacles for routing during a predefined period. Automated calculations of routes have to take into account these temporal barriers, especially furnished with attributes for handicapped.

Fig. 1. Object inventory of the street coverage

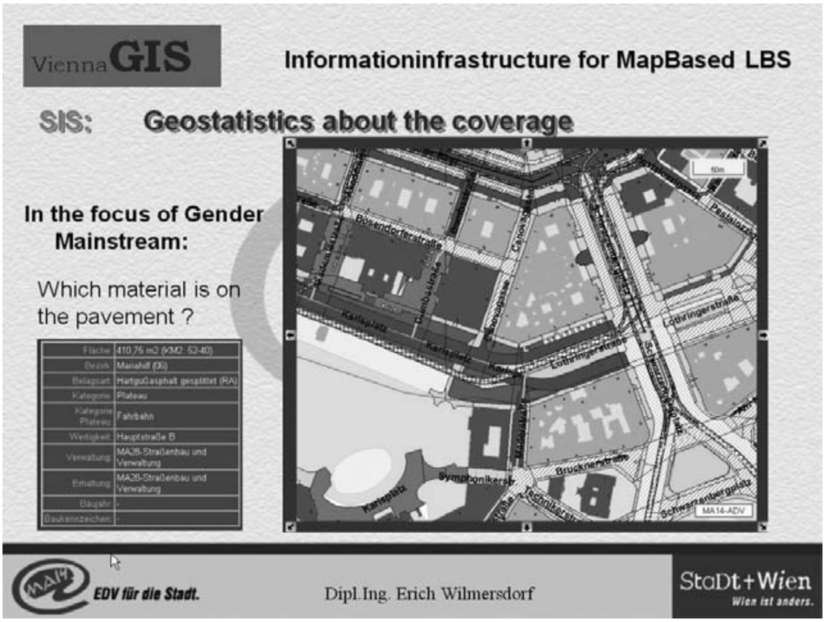


Fig. 2. Geostatistics in the Intranet: material of the pavement

Despite the considerable scope of Vienna´s already existing Geodata-Warehouse there are still deficiencies; e.g. additionally to the public street network private streets, foot paths and passages for pedestrians, temporal obstacles on the side walk (open air events) are needed for improving automated routing recommendations and consequently for LBS purposes. The “GIS-Masterplan” of the City of Vienna - passed 2004 - is a development scheme for completing the urban GIS. It envisages the complete penetration of the administration with GIS-facilities: It focuses on the gaps in the recent “ViennaGIS GeodataWarehouse” to be filled until 2007/2008. This encouraging goal will be essential for the efficiency of administrative processes but also for extended LBS usage in terms of completeness and topicality.

3.3 Change Management of Geodata

Successful LBS afford an utmost topicality for reliable analysis. Therefore the maintenance of the data must be organized in an efficient way.

3.3.1 Utmost efforts for quick updating

To cope with the demand for utmost topicality changes in the urban area have to be recorded digitally as soon as possible. Digital recording facilities are placed in those branches of the administration, where changes are planned e.g. public projects (Wilmersdorf, 2003) and where changes are observed at first. Especially the integration of GIS into administrative processes is expected to support these efforts: Data are collected in the daily workflow. Early bird information about pending changes can be recorded in advance by tapping the data flow of processes, furnished with the date of becoming effective.

3.3.2 Sophisticated change management

Many ways of updating are applied: Electronic measuring techniques in the field, CAD input imported directly from private contractors, GIS based editing in house. But also a WebEditor offers functions for recording geo-related data on each PC in the urban telecommunication network without a GIS software licence for simple geometric input together with object`s identification keys, timely aspects and attributes So the urban administration of Vienna is able to take charge of updating geo-data on those working places, where new data occur at first. A mobile version of this Web-Editor is aimed to support directly the collection of new data in the field. All these methods are applied in order to solve updating in an economic and quick way as early as possible. Finally all these efforts together should guarantee that the administration of Vienna contributes its part to the topicality of its Geodata-Warehouse, as a sound basis for LBS.

3.3.3 Joint efforts for data maintenance

It is obvious that the information infrastructure of the public administration is only one part. Data bases for privately run services must be updated by the private enterprise as fast as possible likewise. User oriented information about the kind of service and any restriction must be offered in time and in a reliable manner, so that LBS users get confidence into the service, an important prerequisite for acceptance. In cases of different data sources a corporate approach for data capture within a LBS application is absolutely necessary.

Cooperation with IT-networks promotes the integration of geographic data with service oriented attribute data. It splits up the amount of work for updating as well. Therefore the City of Vienna has signed cooperative contracts with external authorities and institutions for mutual data exchange; e.g. data about the physicians are maintained by the Chamber of Physicians and they are incorporated into the urban geo-data warehouse for internal use but also for external services in case of emergency.

4 Rendering Geographic Services

4.1 Analytical features

LBS are faced by individual requests for points or areas of interest. So a flexible combination of themes and quick response by assembling the adequate information package is claimed. Analysis services executed by the GIS Server are offered by the city: spatial retrieval in combination with selected attributes and object classes e.g. a 3D-analysis of the terrain model together with the street network to find a route for a bicyclist which does not exceed a maximum limit of a certain gradient.

4.2 Real Time Cartography

To provide the utmost topicality of information, the building up of the cartographic model must be possible on call, exploiting the current state of the databases. Dynamic maps by real time processing are the adequate answer offered in the Internet services of the City of Vienna.

Dynamic compilation follows basically according to this guideline: Map modelling is generally launched on call generating cartographic objects derived from the original object data sets. The cartographic compilation is executed on basis of visualisation rules for each object class and attribute. There are many kinds of cartographic presentation, depending on the scale, the level of detail and the set of themes. Many aspects have to be covered:

- visualisation has to be adjusted to fit the individual demand for geographic information, so to speak “maps made to measure” (e.g. scale dependent construction according to interactively defined space and content, accompanied by automated text placing)
- visualisation has to fit the requirements of the output device (size and resolution of the display).
- Additional information offered in a dialogue by clicking on an object in the map to get more information about the object in a separate window.

These functions are still in the focus of research, to improve the automated map processing for LBS-presentations in cooperation with the user interface.

4.3 Providing Web Services

4.3.1 Infrastructure

The challenge: Individual questions about services of the City have to be answered customer related. Geographic Web services created a new level of accessibility for geo-related information (Groot & McLaughlin, 2000), marked by adaptable map services. But these map based Web Services claim computing power. A solid infrastructure of servers is necessary to cope with the workload 7 x 24 hours a week. Processing is distributed to different servers executing special tasks in parallel mode as far as possible: Checking input especially the local address, interpreting the theme asked for, address localisation, calculation of the map extent, selection of the themes, constructing the map image. There are three levels of Web services, which are discussed in the following section:

4.3.2 Map-based Internet Services

Prefabricated service rendering applications – backed by GIS functions for identifying locally - offer information processing guided in a dialogue. All Internet services of the City of Vienna with geographic aspects are based on:

- real time cartography
- common Browsers (no Plug In)
- customer´s interaction
- flexibility in conveying geo-related information in different ways

Individual navigation and geographic analysis allow the creation of flexible maps on the display e.g. the Viennese internet services: “Stadtplan mit Adressensuche”, also available for PDAs and the “Cultural Inheritance Information Service” (see: http://www.wien.gv.at/wiengrafik) The last one has been incorporated with other external cultural databases in a pilot project “LBSCult”.

4.3.3 WebMap Services (WMS)

The municipality supports also private projects for geographic information services. They are handled on a subscription basis, the selection is defined by parameters. Map images can be transferred according to the user`s preferences. Standards according to the Open GIS Consortium (OGC) are applied to support the transfer of cartographic data on call to external customers. Via a request over the client the spatial identification key and selection criteria defined in the parameter list are transferred to the Web Server, the spatial extent is calculated and according to the parameters the map image is constructed in a tailored way to customer’s need. The platform independent map images (raster data) can be integrated into applications of the customer and put together with his own geo-data.

4.3.4 WebFeature Service (WFS)

This kind of Web service, based on OGC Standards, allows the download of vector data according to the customer’s request. Certain object classes are selected within the defined space. Identification and selection routines generate an individual package of data transferred into the map projection asked for and converted into the agreed data exchange format.

5 Conclusion

LBS need a sophisticated provision with geo-information on a broad scope of detailed and current space-related data (Ladstätter, 2002). Each single enterprise is unable to maintain such an infrastructure completely by its own. Cooperation is necessary. Municipalities are important players in this context. They have to run voluminous knowledge data bases geo-related in order to manage their tasks, to monitor and control the development of the urban region. Their knowledge data base represents a valuable source for thematic and cartographic data. So municipalities have the prerequisites and therefore the commitment to play a key role as an information dispensing enterprise. Their IT-infrastructure for their internal processes can also be exploited for geographic services for the public, even to fit the special requirements of map based LBS. Municipalities are expected to act not only as distributor of geo data but also as provider of attribute data and of value-added information needed by LBS. Most important is the strict commitment to guarantee update as fast as possible.

A municipality`s contribution in the field of LBS can be seen as twofold:

- Providing own LBS offering public services: The city of Vienna runs a broad scope of Internet services, which could be presented as mobile services for the public by LBS as well.
- Establishing an infrastructure for private enterprises with links to the urban data bases (offering raw data but also value-added information).

The Municipality of Vienna has declared it as a strategic goal to act as a service providing enterprise. LBS are one of the many ways to enhance its service level. So it promotes the effective exploitation of its Geodata-Warehouse inside but also from outside. Information should not be available in the body of the administration only, information has to be transferred to the place, where and when it is needed. Committed to standards access is facilitated for new LBS-applications. Future inhouse-developments are focussed on generating software increasing the flexibility of dialogue-oriented Web Services and especially of real time cartography together with multi-media to improve the presentation of information.

References

Bartelme, N. (2002) GIS Technology, Springer Verlag

Groot, R., & McLaughlin, J. (Eds). (2002) Geospatial data infrastructure: concepts, cases and good practice. Oxford University Press

Ladstätter,P. (2002) TeleKartographie &Location Based Services, Geowissenschaftliche Mitteilungen (Heft 58) TU Wien

Wilmersdorf, E.(2003) Geocoded information incorporated into urban online services – the approach of the City of Vienna, Elsevier Science Ltd: Computers, Environment and Urban Systems

Wilmersdorf, E. (2003) Ein Fall für eGeogovernment-Aufgrabungen im Straßenbereich und zentraler Leitungskataster, in geoGovernment, Verlag Wichmann

MoGeo: A location-based educational service

David A. Bennett, Marc P. Armstrong, Jerry D. Mount

Department of Geography, The University of Iowa, Iowa City, USA
david-bennett@uiowa.edu

Abstract

The classroom is an efficient venue for conveying theories and concepts in many disciplines. In geography, however, students often have difficulty understanding spatio-temporal processes that are represented abstractly. Though field-based classes embed students in the environment they are studying, they are decoupled from supplemental information that often catalyzes learning. In this paper we present a location-based service for geographic education: the Mobile Geographic Education (*MoGeo*) System. *MoGeo* design is based on contextually-aware computing and integrates the following technologies: GPS receivers, GIS software, wireless networks, mobile computers (PDAs or tablets), and centralized computers (*e.g.*, servers). Using this assemblage of technologies, we integrate the important elements of the classroom and computer laboratory into a field-based learning environment by providing remote, real-time access to: 1) context-specific educational materials; 2) sophisticated spatial analyses; 3) high-end visualization and simulations; 4) feedback and evaluation; and 5) instructors and peers. The overall experience is analogous to the guided tours available at many museums, but is richer and more conducive to intellectual exploration (*e.g.,* through simulation, visualization, and experimentation) and collaboration (*e.g.,* through audio/visual communication and real time tracking of classmates).

1 Introduction

The purpose of our ongoing research project is to promote geographic understanding through experiential, field-based learning opportunities supported and enhanced by location-based educational service technologies. The use of these technologies allows students to take the classroom, computer laboratory, library, and Internet with them into the field. By integrating the best elements of these more traditional learning opportunities, complex and abstract spatial concepts come alive and assume greater real-world meaning for students (Armstrong and Bennett 2005). This educational experience is referred to as Mobile Geographic Education (*MoGeo*). In the *MoGeo* framework, “where” is both paramount and, paradoxically, irrelevant. The delivery of place-specific, highly contextualized information about what students are to learn given their current location on the landscape, is what makes *MoGeo* unique and powerful—the “where” matters when learning about geographic processes. On the other hand, recent advances in mobile computing technology provide access to instructors, peers, knowledge repositories, data, and computational power from almost any location—“where” no longer matters when it comes to the provision of educational materials. These characteristics of *MoGeo* have challenged us to rethink the way knowledge is transmitted as we disconnect educational processes from traditional classroom settings and embed them into the system under study.

To understand the conceptual framework on which *MoGeo* is built, it is useful to make a comparison between the production of knowledge through scientific inquiry and its dissemination in higher education. First, consider how knowledge is produced (Figure 1). Reality is observed and measured. To better understand what we empirically observe we develop theories about how reality works. Experiments and models are constructed to help us examine the validity of our theories in manageable, simplified settings; experiments and models provide a bridge from theory to reality. What we learn from these experiments and models helps us rethink our theories, and what we learn about our theories can change the way we view reality. Now consider how these three elements of knowing (empirical observation, theory development, and experimentation/model building) are translated into three common teaching environments, the classroom, laboratory (here we constrain discussion to the computer laboratory), and field class.

For centuries, the classroom has proven to be an efficient, if imperfect, mechanism through which theories and concepts can be conveyed. Although the classroom provides a forum for the presentation and discussion of ideas, the representation of many spatio-temporal phenomena of interest to geographers is often restricted to highly abstract diagrams or static images of reality because of the scale of observed processes. In laboratories students get a “hands-on” feel for the interrelationships that drive dynamic spatial processes through space and time using visualization and simulation software. Furthermore, students can augment their knowledge through the review of related resources that are accessible from the Internet. While there is no consensus on whether or not Internet sources improve learning, students tend to place value on multiple data sources that present information in alternative forms (Chrisman and Harvey 1998; Ritter and Lemke 2000; Jain and Getis 2003). However, two problems often constrain the learning process in computer laboratories. First, traditional laboratories are ill-suited to collaborative learning (Reed and Mitchell 2001) — too often students sit in front of their machine focused on the implementation of complicated computer instructions, taking little opportunity for the kinds of discussions that can occur in classrooms, or group problem solving activities that often occur in the field. Discussion and group problem solving, we maintain, assume importance when students are asked to make a cognitive connection between model and theory. Similarly, the bridge from simplified model to a complex reality can remain tenuous because of the level of abstraction needed to implement models in laboratory settings.

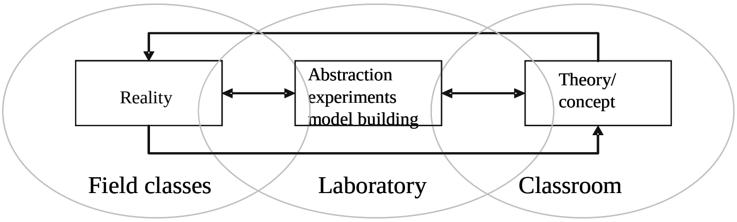


Fig. 1. Knowledge is produced through a highly integrated process of observation, experimentation, and theory development. These three “elements of knowing” are often fractionated when knowledge is disseminated in higher education.

Field-based classes link the student to reality and many studies document the importance placed on field-based experiences by students (see, for example, Kent, Gilbertson and Hunt 1997; Fuller, Rawlinson and Bevan 2000; Pawson and Teather 2002; and Fuller, Gaskin and Scott 2003). However, field-based classes have historically been isolated teaching environments because of: 1) difficulties associated with accessing the kinds of supplemental material that help catalyze learning; and 2) logistics, if students become too dispersed they cannot effectively interact with the instructors or each other.

The highly integrated process of knowing, therefore, becomes transformed into a decidedly more fractionalized process of teaching. *MoGeo* is intended to overcome this difficulty by fusing the important elements of the class, laboratory, and field into a single learning environment that provides students with *in situ* access to: 1) knowledge and data resources stored in web accessible repositories; 2) state-of-the-science spatial analytical tools; 3) the knowledge of teachers and peers; 4) a way to communicate at a distance; 5) models and visualizations of processes under study; and 6) real-world geographic context. By providing the communication and computation tools needed to take the classroom and the computer laboratory to the field, *MoGeo* promotes scientific discovery and, we maintain, the dissemination of knowledge becomes more closely bound to the production of knowledge.

2 Principles to guide the application of *MoGeo* in higher education

Ten prescriptive principles for *MoGeo* application design were presented in an earlier paper (Armstrong and Bennett 2005). While we do not expect that all projects will strictly adhere to every principle, the relevancy of each should be considered as one designs *MoGeo* enhanced laboratory exercises. These principles are summarized as follows (for greater detail please refer to Armstrong and Bennett 2005):

1. Promote *in situ* learning experiences - Directly couple *in situ* field experience to data and knowledge repositories accessible via the World Wide Web to help students contextualize abstract concepts and explore related ideas.
2. Use spatial triggers (intelligent landmarks) - Turn the field experience into a self-guided tour analogous to those available at many museums by monitoring students' positions and delivering context-specific information at appropriate times and locations.
3. Accommodate multiple learning styles - Support alternative learning styles by producing class materials in graphical, textual, auditory, and symbolic (mathematical) form (Ritter and Lemke 2000; Smith 2002).
4. Produce interactive, dynamic, and student-centered learning experiences- Engage students in the learning process by promoting hands-on problem solving activities and context specific interaction.

5. Teach about the importance of spatial relationships and their digital representation- Produce laboratory exercises that require students to conceptualize and implement strategies for the representation and capture of attribute, geometrical, and topological data.
6. Teach about proper editing practices and the importance of metadata- Promote the proper and confident use of geographical data by requiring the development and maintenance of metadata that traces the history of alterations made by data users.
7. Teach about privacy and the ethical use of *MoGeo* technologies- Promote the ethical use of mobile computing (Armstrong 2002; Monmonier 2002; Beresford and Stajano 2003; Myles, Friday and Davis 2003; Gruteser et al. 2004) by illustrating the responsible use of related technologies (CSTB 2003).
8. Promote personal safety- Make sure that laboratory exercises do not expose students to danger or cause them to behave in unlawful ways.
9. Promote the safe and secure use of wireless technologies- Be sure that the use of wireless technologies in laboratory exercises does not to expose campus computer systems to unauthorized use or malicious attacks.
10. Develop efficient code for mobile devices- When developing laboratory exercises always consider the limited computing resources associated with mobile computing technologies (Viredaz, Brakmo, and Hamburgen 2004) and avoid the frustrations associated with slow or inconsistent technologies.

When considered collectively, three overarching themes emerge from these principles. First, context is important. We must know who the students are and the environment in which they are working. Second, the system must be flexible. It should be possible to easily modify a *MoGeo* learning environment to, for example, represent a new landmark or support alternative learning styles. The final theme is responsibility. The potential for abuse with mobile technologies, by both instructor and student, is clear and steps must be taken to insure that all parties understand their responsibility for the ethical use of these devices.

3 *MoGeo* Technology

The design of *MoGeo* is based on the concept of contextually-aware computing. The system serves educational material to a student when and where it is needed to complement empirical observations and knowledge gleaned from classroom discussions. To accomplish this coordinated push and pull of information we must know who students are and the class in which they are enrolled. We must also know the laboratory exercise particular students are attempting to complete and where they are located in the landscape. Given this information we can match the students' context (*i.e.*, location and laboratory assignment) to a set of spatial triggers stored in the system. Each landmark is associated with a spatial feature we want students to learn about (*e.g.*, a wetland or particular land use) or locations at which we want them to perform some task (*e.g.*, collect data or modify a spatial database). Persistent agents, instantiated on the server, monitor student context, match student context to landmark characteristics, and, if appropriate, coordinate the push and pull of context-specific information. Using contextually triggered "pop quizzes", we are able to assess how well students meet learning objectives, identify where mistakes were made (*e.g.*, discrepancies between the students work and expected product), and take remedial actions while the student is actively engaged in the learning process. The student could, for example, be directed to recapture points that were incorrectly located, complete a step that was inadvertently skipped, search web-based resources that describe a particular concept in greater detail, or "get a call" from the instructor using voice over internet protocol (VOIP).

The *MoGeo* system consists of the following components: 1) mobile computing devices; 2) GPS receivers; 3) WiFi wireless communication cards; 3) GIS software; 5) a notebook or desktop-based server; 6) wireless communication access points; and 7) communication software. These components are deployable in two configurations. The server can be attached directly to the hardwired campus communication network and linked to the mobile devices via outdoor wireless access points installed on campus buildings. This configuration provides users with full access to network accessible knowledge repositories, but restricts the spatial range of the system to those areas covered by installed access points—an area less then $1km^2$ given current technology. A field-based configuration can also be deployed. In this situation, a wireless access point and directional antenna is attached to a notebook computer. This configuration frees the system from the campus backbone and, thus, allows it to be used at almost any location. If an Internet connection is unavailable for the notebook system, however, a knowledge repository must be developed and installed on it before going to the field. This limited connectivity, of course, reduces the total volume of information available to students. With WiMax and the proliferation of web-enabled cellular technologies, however, connecting to the Internet should not be a long term impediment to *MoGeo*.

Figure 2 illustrates the interactions among various *MoGeo* software and hardware components. Students log into the system, linking their digital identity to a class, laboratory exercise, and location stream (the stream of points produced as a student moves across the landscape). The GPS samples the students location at regular intervals (*e.g.*, every three seconds) and software loaded on the PDA sends a "who, what, when, and where" information packet to the server through the wireless communication link. In our current system, the software used to capture, store and query data on the server side is SQL Server. Controlling logic written in XML queries these data to determine if students are within a trigger space associated with their current laboratory exercise. What happens when a trigger event occurs is, of course, dependent on the laboratory exercise. PDAs and tablet computers are the two most practical computing form factors available at this time for supporting *MoGeo*. At roughly one third the cost of a tablet, PDAs are clearly the most cost effective solution, but they have two significant limitations. First, with a display area of approximately $50cm^2$; these devices have a limited ability to present the kinds of cartographic representations that GIS users have come to expect. Second, these devices have limited computing power and, thus, cannot perform computationally intensive geographical analyses. To overcome this problem requests for such analyses can be uploaded to compute servers and the results sent back to the mobile device. The implementation of such a solution, however, often requires custom software development.

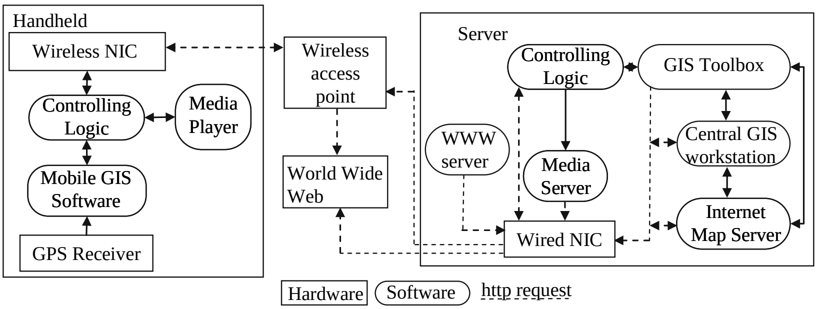


Fig. 2. Schematic of *MoGeo* implementation

We have been developing *MoGeo*-style capabilities for about two years. We currently deploy 20 PDAs (Compaq IPAQ model 3850, purchased in 2002) each of which runs a mobile GIS software package (ESRI ArcPad) and software developed "in house" to support communication among components and the transformation of incoming data to a GIS format (Figure 3). These devices are equipped with WAAS compatible GPS units and WiFi network interface cards. The compute server is a 3.4 GHz

Pentium 4 notebook computer with 1GB of RAM and a 100 GB disk. This server is equipped with a WiFi network interface card, GIS software (ESRI's ArcMap 9.1, associated extensions and ArcObjects), and SQL Server. Persistent agents running on the server provide two main location-based services for the user. First, they collect incoming data from mobile devices and coordinate the push and pull of context specific information. Second, these agents coordinate analytical tasks that are beyond the capability of the mobile GIS software or the PDA. These tasks are executed using calls to ArcObjects.

Fig. 3. A ***MoGeo*** device consists of a mobile computer (handheld or tablet), a global positioning system receiver, wireless access, GIS software, and communication software.

4 Teaching the four traditions of geography with *MoGeo*

In 1964 William Pattison established four themes that he believed captured the intellectual core of geographic enquiry (Pattison 1964). These "four traditions of geography" are: 1) spatial; 2) area studies; 3) man/land (what we now refer to as human/environment); and 4) earth science. Twelve years latter, Robinson (1976) found that these traditions retained their relevance to the discipline. While some may debate whether additional subdisciplines should be added to this list (we can, after all, forgive Robinson and, in particular, Pattison for not foreseeing the importance of GIScience in the days of the mainframe, key punch, and line printer), this set remains relevant today and is sufficiently diverse to illustrate the range of educational applications to which *MoGeo* can be applied. We are in the process of creating *MoGeo* enhanced laboratory exercises that teaches concepts associated with each of the four traditions of geography. These exercises are described below.

4.1 The spatial tradition

In discussing the spatial tradition of geography, Pattison and Robinson remind us of the central role that geographic distributions play in our discipline. The distribution of geographic phenomena both defines and is defined by spatial interaction and the movement of, for example, goods and people across geographic space. How do we make tangible the concepts of impedance, connectivity, and accessibility that are typically dealt with in the context of abstract graph-based routing algorithms designed to operate on large scale geographical problems? How do we change the scale of the problem so that it can be addressed in an academic setting while maintaining the salient aspects of real-world problem solving? Our solutions to these questions are inspired by the project "Modelling Access with GIS in Urban Systems (MAGUS)" (Mathews *et al*. 2003). Mathews *et al*. (2003) evaluated how urban infrastructure impacts the movement of individuals who use wheelchairs. Wheelchair users were surveyed to determine the relative magnitude of various barriers (*e.g.*, stairs and curbs) and experiments were conducted to evaluate the impedance value associated with alternative surfaces (*e.g.*, cobblestone vs. pavement). Using these data they produced interactive GIS-based datasets that helped wheelchair users navigate within a relatively compact urban shopping area. The laudable goal and useful results of this project can be transferred to any campus.

In this laboratory we will ask students to see the campus from the mostly unfamiliar perspective of one who uses a wheelchair to facilitate movement. From this experience we hope that they will gain insight into how access controls spatial movement and, from a more human perspective, insight into the lives of those who might be different from them. The specific objectives for this laboratory are for students to learn:

1. field-based data collection techniques,
2. the integration of collected data into an existing dataset,
3. how to develop linear referencing systems,
4. alternative distance metrics,
5. the theory and application of routing algorithms, and
6. the importance of accurate and up-to-date information.

To make the problem tractable in the context of a laboratory exercise, students are provided with a partially developed database that they must verify, update, and extend. This database can store street crossings, wheelchair accessible building entrances, road and sidewalk surface types and other impediments to movements (*e.g.*, steps) that exist on campus. Locational triggers fire when a student approaches points or segments that must be verified or added to the database. By integrating the “as is” data with digitally produced (virtual) alternatives, we can add complicating factors into the problem. For example, as they enter one trigger area they will be informed that where they stand is, for the purposes of the exercise, under construction and, thus, the surface is temporarily gravel (not the concrete that is under their shoes). A digitally altered and geographically referenced orthophotograph pops-up on their *MoGeo* device so that they can map the bounds of the construction project. Point features and surface data would then be integrated into an existing sidewalk dataset of the campus. Using data documented in Matthews *et al.* (2003), network attributes (*e.g.*, turning movements through points, impedance values along segments) would be edited to produce a transportation model of the campus from the perspective of wheelchair users. The students then return to the field and, using the original and modified network datasets, produce and traverse shortest paths using GIS software. Using the spatial location provided by the GPS, and time provided by system clocks, we can track the students and enforce the impedance values associated with wheelchair movement.

The students’ task will be to construct and traverse least cost paths using networks based on walking and on the use of a wheelchair. As students move between assigned locations they have to climb gradients that are too steep to negotiate in a wheel chair—impedance to movement can be anisotropic (Figure 4). The steep slope requires students to travel out of their way to use elevators in buildings that have access points at both the top and bottom of the gradient. The compute server can be used to calculate alternative paths in real-time (Figure 5). To drive home the importance of timely information, we can also produce virtual construction projects that block what would otherwise be shortest paths—requiring the student to find alternative and, thus, more taxing routes.

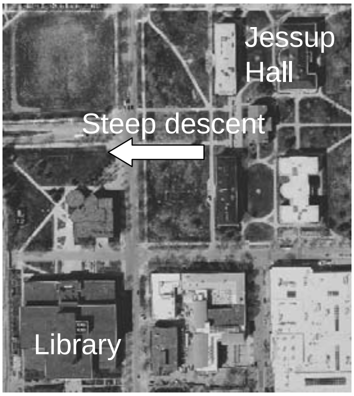


Fig. 4. Students produce spatial datasets that document impediments to movement for those who require wheelchairs to move about campus. These datasets will then be used to produce shortest path routes between campus facilities. From this exercise students will gain insight into how access controls spatial movement and, from a more human perspective, insight into the lives of those who might be different from them.

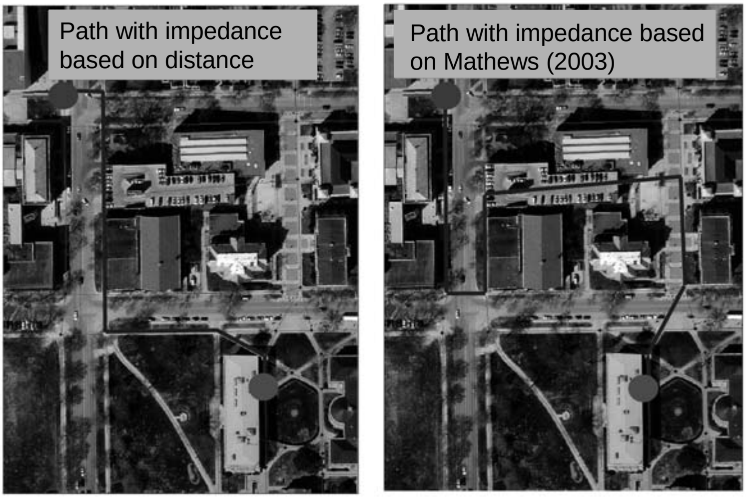


Fig. 5. Students learn how human perception of geographic space is affected by impediments to travel and about the mechanics of automated routing systems by developing a real time navigation system for the mobility impaired.

4.2 The area studies tradition

The second tradition explored by Pattison and Robinson is area studies. Area studies are concerned with the characterization of places, areas and regions—what makes an area identifiably different from another. In our classes we talk about how geographers define a region, but it is difficult to convey the subtlety, complexity, and ambiguity of the regionalization process in the context of a lecture. What if students had to regionalize an area around a campus? What typologies would they develop and what criteria would they use to distinguish among the classes that make up these typologies? Students in the field equipped with *MoGeo* technology will have the ability to analyze and map, for example, census data, zoning maps, and transportation facilities. These data can be integrated with geographically referenced aerial photographs to provide detail and render abstract socio-demographic data more concrete. The question becomes: Can they place these data into context and extract the more intangible qualities that define a region using direct observation?

Through this exercise we expect students to gain an appreciation for the:

1. concept of geographic regions and the difficulties associated with indeterminate boundaries,
2. challenges of semi-structured problem solving,
3. ethical use of MoGeo, and
4. strengths and limitations of socio-demographic data.

Students can be organized into teams and sent to the field to develop general impressions of the selected study area. Each team develops a set of hypotheses about what they believe explains the spatial patterns that they see in the field. Considerations might, for example, include zoning, percent of owner occupied housing, median household income, and housing value. Teams then return to the field, analyze available data *in situ* and (re)evaluate their hypotheses. Students can construct thematic maps of socio-economic data stored in the department's database (or downloaded from the Census Bureau if Internet access is available) and overlay these maps onto the street map or aerial photograph to provide context. Using the compute server, statistical surfaces can be produced from numerical variables and the student can visualize transitional zones that might indicate boundaries between regions. The students can use VOIP to contact their instructor if they have questions about specific kinds of data or need clarification on project objectives, or other student groups if they want to compare what they see at their location to what others see elsewhere; communication could be enhanced with the ability to capture and send digital images. The location of all individuals and teams can, of course, be displayed on the *MoGeo* devices (*i.e.*, everyone will see where everyone else is). This will facilitate communication among team members or among different teams (*e.g.*, I see that you are in block group 10 and it has a median income of $40,000, how is it different then block group 11 that we surveyed together yesterday? Please send images).

Location triggers will be used to highlight important issues. For example, we can address:

1. concerns about the use of individual-based data and the balance between the public's right to know and individual rights to privacy (*e.g.*, the practice of mapping the location of those accused of specific crimes).
2. the difficulty of drawing sharp cartographic lines to represent transitional phenomena.
3. issues associated with error and uncertainty.

4.3 The human/environment tradition

The human/environment tradition addresses the reciprocal relation between humans and the natural environment in which they live. Whereas geographical research during the first half of the 20th century considered how the environment shaped humans, work during the second half of the century began to focus more on how humans reshape the environment. As humankind attempts to manage the environment, it must balance an array of often competing objectives. This balancing act became evident to the residents of Iowa City, IA during the summer of 1993 when the Midwest was confronted by widespread flooding. We use this experience to motivate our analysis of human/environment interactions.

The University of Iowa sits on the banks of the Iowa River just downstream from the Coralville Reservoir. This impoundment is managed by the U.S. Army Corps of Engineers for flood protection and recreation. During the flood of '93, Corps personnel struggled to balance inflow and outflow from the reservoir. If too much water was released, flood conditions in Iowa City would worsen (conditions that already threatened university buildings and water supply), if not enough water was released, the reservoir would no longer be able to absorb future rains. In this laboratory exercise a simple systems model will be used to simulate the 1993. Our objective is to teach students about:

1. our ability to manage natural processes,
2. simple modeling of natural processes, and
3. links between policy, development, and natural processes.

The students will begin the laboratory exercise next to the river learning about the flood. They will connect to the USGS real-time river monitoring web-page and gain an appreciation for the flow conditions they are observing. Using stage-discharge relationships, a map will be produced that shows the extent and depth of flood water in the area that they are standing, given alternative assumptions about river discharge. Flood video from the area will be delivered to the students so they can see 1993 conditions. As students tour the campus, spatial triggers will fire and context-relevant information will be delivered describing, for example, the: 1) threat to the university water treatment plant; 2) damage caused to infrastructure and property; and 3) impact on transportation. Web links to such regional stories as the disruption of water supply in the city of Des Moines, IA during the ’93 flood and the more recent buy-out programs for flood damaged homes in northeast Iowa will stress the widespread and ongoing impact of flooding in Iowa. Students will be provided Internet links to the federal flood insurance program and asked to consider how policy might affect floodplain development.

Next we will ask students to play the role of Corps personnel. Using a simple reservoir model they will adjust discharge rates “coming from” the reservoir. This system will be subjected to a series of hypothetical rainstorm events that will determine the inflow rate to the reservoir and the students will see on their mobile devices the relation between inflow, discharge rate, reservoir level and flooding. Because these events are artificial, we can monitor the actions that each student is taking to produce desired outcomes (e.g., downstream flooding, low flow, dangerously high reservoir levels). The intent is to engage them in a “SimRiver” experience with the simulated results mapped out on their *MoGeo* device and the actual landscape in front of them for perspective.

4.4 The earth science tradition

Scholarly work in the earth science tradition explains how continuous biological and physical processes operate across space and though time. To study these processes students must possess a solid understanding of sampling and interpolation techniques. In this laboratory exercise students will use these techniques to construct digital representations of topography (*i.e.*, a digital elevation model). Through this exercise students will learn:

1. the effects of alternative sampling strategies on interpolation results
2. the effects of alternative interpolation techniques on results
3. techniques to assess accuracy, and
4. the limitations of data collection devices.

The study area to be used for this laboratory will be carefully surveyed using a total-station (a high accuracy survey instrument) and these data will be used to produce an accurate digital elevation model (DEM). Using locational triggers, students are directed to predetermined locations within this area. Each student will capture his or her coordinates in 3D-space using their GPS (collectively, they constitute a sampling network). These coordinates will be sent to the server, which will use these points to interpolate a DEM. A difference map can then be produced to calculate and visualize the spatial pattern of error by subtracting the student produced DEM from the more accurate total-station produced DEM. This difference map will be sent back to the students. By keeping the students stationary and repeating the process several times the effect of measurement error and variability on the data collection process can be illustrated. The elevation values provided by our moderately priced GPS units possess too much error for the construction of accurate DEMs (a fact that should become obvious to students as they analyze the difference maps they produce). To overcome this problem, we can use the horizontal position from the GPS to extract elevation values from the high accuracy DEM and use these values in the interpolation algorithms.

The *MoGeo* devices can then be used to direct students into alternative sampling configurations (*e.g.*, regular, random, stratified random, adaptive), new sets of points captured and new error maps produced. With each new configuration students will be encouraged to discuss the relation between the location of sample points (evident by the location of their fellow classmates) and error. Given this experience, students are then asked to produce a DEM with minimal error by adapting their sampling strategy (*i.e.*, moving their positions in space). It is expected that through this process of experiential learning students will discover that error is reduced when, for example, valleys, ridges, and slope breaks are captured. Similar procedures can be used to investigate the impact that the number of sampling points has on interpolation results.

5 Future work

We have only just begun to explore the possibilities of *MoGeo* and further work is needed on all aspects of this endeavor. We need, for example, to:

1. rigorously evaluate the impact of *MoGeo* on the learning process;
2. explore the utility of multiple communication channels (i.e., voice, video, data);

3. evaluate the impact of potential technological constraints, such as bandwidth, response time, and distance, on user satisfaction;
4. design, implement, and evaluate human-computer interfaces that facilitate the *MoGeo* learning process;
5. design, implement, and evaluate cartographic and geovisualization products well suited to the *MeoGeo* learning process; and
6. evaluate the utility of real-time feedback to students during the learning process.

6 Conclusions

Locationally enabled mobile computing is beginning to routinely affect our daily lives. As a consequence, it is important that geography students, in general, and GIScience students, in particular, understand the strengths and weaknesses of emerging technologies by learning to use and apply them properly. Such understanding and hands-on experiences will give our students the knowledge needed to: 1) compete successfully in a rapidly changing technological world; and 2) use advanced geospatial and mobile technologies in an appropriate and ethical manner. We further contend that the need for such an understanding is, in fact, widespread within the academy and that the activities described in this paper will have important ramifications on a wide range of disciplines, literally from Anthropology (e.g., studying the ecological context of ancient sites) to Zoology (e.g., calculating habitat suitability while conducting field surveys

While an educational example is described in this paper, the concepts embedded in *MoGeo* are portable to LBS in general. For example, conceptual maps derived from user interaction can attach "value added" information to landmarks (*e.g.*, individuals who found this attraction interesting, also enjoyed these attractions…), while virtual locks monitored by persistent assistants can filter out unwanted solicitations (*e.g.*, forward only those services associated with landmarks containing historical information). The opportunities seem endless. The trends we observe today in mobile, embedded, ubiquitous, and pervasive computing are, after all, just the metaphorical tip of a very large iceberg.

7 References

Armstrong, M.P. 2002. GI technologies and their potentially erosive effects on personal privacy. Studies in the Social Sciences, 27(1):19-28.

Armstrong M.P. and Bennett, D.A. 2005. A manifesto on mobile momputing in geographic education. Professional Geographer, 57(4):506-515.

Beresford, A. and Stajano, F. 2003. Locational privacy in pervasive computing. Pervasive Computing, **2**(1): 46-55.

Chrisman N R. and Harvey, F.J. 1998. Extending the classroom: hypermedia-supported learning. Journal of Geography in Higher Education, 22(1): 11- 18.

CSTB – Computer Science and Telecommunications Board Committee on Intersections Between Geospatial Information and Information Technology. 2003. IT Roadmap to a Geospatial Future. Washington, DC: The National Academies Press.

Fuller, I., Rawlinson, S., and Bevan, R. 2000. Evaluation of student learning experiences in physical geography fieldwork: paddling or pedagogy? Journal of Geography in Higher Education, 24(2):199–215.

Gruteser, M. and Xuan L. 2004. Protecting privacy in continuous location-tracking applications. IEEE Security & Privacy, **2**(2): 28-35.

Jain, C and Getis, A. 2003. The effectiveness of internet-based instruction: an experiment in physical geography. Journal of Geography in Higher Education, 27(2):153–167.

Kent, M., Gilbertson, D.D., Hunt, C.O. 1997. Fieldwork in geography teaching: a critical review of the literature and approaches. Journal of Geography in Higher Education, 21(3): 313- 332.

Matthews, H., Beale, L., Picton, P. and Briggs, D. 2003. Modelling access with GIS in urban systems (MAGUS): capturing the experiences of wheelchair users. Area, 35(1): 34–45.

Monmonier, M. 2002. Spying With Maps. Surveillance Technologies and the Future of Privacy. Chicago, IL: University of Chicago Press

Myles, G., Friday, A. and Davies, N. 2003. Preserving privacy in environments with location-based applications. Pervasive Computing, **2**(1): 56-64.

Pattison, W.D. 1964. The four traditions of geography. Journal of Geography, 63(5): 211-216.

Pawson, E. and Teather E.K. 2002. Geographical Expeditions: assessing the benefits of a student-driven fieldwork method. Journal of Geography in Higher Education, 26(3): 275–289.

Robinson, J.L. 1976. A new look at the four traditions of geography. Journal of Geography, 75(9): 520-530.

Reed, M. and Mitchell B. 2001. Using information technologies for collaborative learning in geography: a case study from Canada. Journal of Geography in Higher Education, 25(3):321–339.

Ritter, M.E. and Lemke, K.A. 2000. Assessing the 'seven principles for good practice in undergraduate education' with internet enhanced education. Journal of Geography in Higher Education, 24(1): 100-108.

Bata Positioning System - A real time tracking system for the world's largest relay race

B.J. Köbben[1], D. Boekestein[2], S.P. Ekkebus[2], P.G. Uithol[2]

[1] International Institute for Geo-Information Science and Earth Observation, PO Box 6, 7500 AA Enschede, The Netherlands
kobben@itc.nl
[2] BPS team, http://www.batalive.nl; info@batalive.nl

Abstract

Modern technology makes it possible to develop applications that determine the position of any object. Using affordable off-the-shelf components combined with internet-based applications, we developed the Bata Positioning System (BPS), a real time system for tracking moving objects. It has a focus on accessibility to the user and system scalability. BPS combines the power and flexibility of satellite positioning, mobile telecommunications and open standards compliant internet applications in a unique way, and was first applied for the Batavierenrace, the world's largest relay race.

We will first briefly introduce the Batavierenrace and then explain how the wish to maintain an overview of this race led to the development of the BPS 1.0 prototype. Its basic requirements and general setup as well as the system architecture will be discussed. It will be explained how further refinement of this architecture led to the current stable version BPS 2.0. We discuss how the BPS 2.0 system principally provides Location Services, but has a great potential to further grow into into a genuine real time Location Based Service (LBS). In the last section, the prototype BPS 3.0 system providing such LBS functionality is presented.

1 The Batavieren relay race

The Batavierenrace [1], 'Bata' for short, is a yearly relay-race with participants mainly from the student population of the Netherlands. The route of the Bata covers over 185 km, divided into 25 stages (17 men's stages and 8 women's stages) varying from 3.4 to 11.9 km. At midnight the race starts in the city of Nijmegen and the participants will run throughout the following night and day through the Eastern Netherlands and neighbouring Germany, to arrive about 18 hours later in Enschede. On average over 300 teams participate in this race, and the 2005 edition's 7825 runners earned it the Guinness-book record of 'largest relay race in the world'. For the majority of the participants, the race is just leisure. But there is also a very serious University Competition, in which the best runners battle for the fastest total time.

The history of the race goes back to 1972, when a couple of students from Nijmegen University came back from a similar relay run in Sweden determined to organise such an event in the Netherlands. The resulting first race was run from Nijmegen to Rotterdam, based on the route of the *Batavieren* tribe, which (according to folk myth) came rowing down the river Rhine some 50 years BC. The Batavierenrace still owes its name to that first route, although it was later redirected towards Enschede.

Most teams use minibuses for the transport between check points, where they have to arrive in time to collect the incoming runner and drop off the participant for the next stage. With all these people moving around and the race taking place mostly in a rural area, partly during the night, it is very hard to be aware of the status of the race. Most of the time, the teams themselves as well as other people interested have no idea where the various runners are at any given moment. Having actually experienced these problems themselves, a couple of students started thinking of a solution that could provide participants and spectators with the location of runners in real time during the race. This was the start of the Bata Positioning System (BPS).

2 BPS: The Bata Positioning System

The development of the system is an ongoing effort by the core BPS team: Three students from the *University of Twente* Telematics BSc/MSc and one from the *Hogeschool of Utrecht* electronic engineering Professional Master programme. This team has been supported by sponsors and several academic supervisors, while some students intertwined this project with their studies, resulting eg. in bachelor theses [2] and [3].

A first version of the BPS was tested in the April 2003 race, and in the following years most parts of the system have been upgraded and refined. In the next paragraphs, the general setup of the current system, which is based on the use of raster geodata, will be presented. Next, there will be a discussion of the next generation BPS, based on vector geodata, and its potential to grow from a Locating Service into a system supporting full–fledged Location Based Services.

3 Basic requirements and setup

As stated in the introduction, the BPS was designed to track runners during a relay-race. Therefore, the prime requirement is that it can provide real time position information of these participants. This information must be presented to the end-user in such a way that the user is able to get an overview of the runner's locations in (near) real time. The system itself must also be easily adaptable and deployable, able to use existing networks and infrastructure. In this way, it can be easily used in any area, for various types of races, without the need for extra infrastructure.

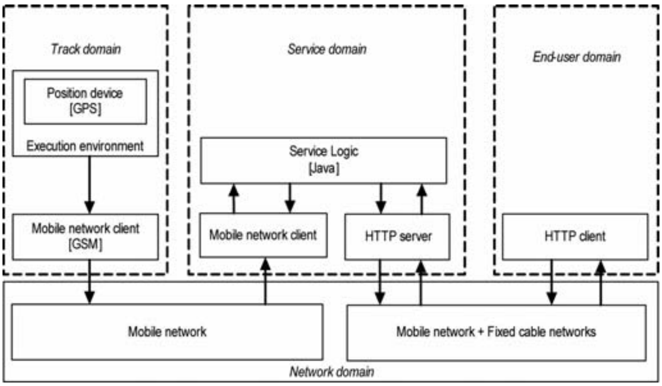


Fig. 1. The four domains of the Bata Positioning System

Four different domains can be distinguished in the BPS setup, as depicted in figure 1. First there is the *track domain*, which provides positioning information about the tracked objects. Each of these tracked objects is an autonomous entity, which uses a push mechanism to transmit its position information to the service domain. The number of objects being tracked by the system must be easily expandable, so the procedure of adding and removing tracked objects is implemented using a plug-and-play mechanism. The positioning device is a simple GPS receiver, which is used by the execution environment to obtain information about its latitude and longitude (on the WGS84 datum) and speed and direction. After the position has been retrieved and some metadata (time, team, runner, etc.) has been added, the execution environment sends the data to the service domain through a mobile network client using the GSM network. Since the existing GSM infrastructure can be considered standardized throughout Europe, this method for transferring data is generally applicable. To transfer the data either SMS, WAP, GPRS or nowadays UMTS can be utilized.

The *service domain* is the central part of the system. Here, the information received from the track domain is parsed by the mobile network client, for use by the service logic, which is implemented in Java. An HTTP server makes the output of the service logic available on the internet.

The HTTP client in the *end-user domain* retrieves the information from the internet and represents the interface to the end-user.

The *network domain* provides the connectivity for the data communication between the domains. Whereas the track domain and the service domain always use the GSM mobile network to communicate, the service domain can be connected to the end-user domain through either a fixed network (dial-in modem, cable-modem, LAN) or any of a number of mobile networks (eg. WAP, GPRS, UMTS, WiFi).

4 BPS 1.0 – the protoype

A first version of the BPS was tested in the April 2003 race, as a limited pilot, tracking only three participating teams, plus the BPS-team minibus. In this first incarnation the *BPS-boxes* (tracking devices) were rather large and filled with the components shown in figure 2: a commercially available GPS receiver module (1), a standard GSM phone (3), connected through and controlled by a self-built microcontroller (2). Not shown in this picture is the battery power pack. Despite testing beforehand, these power supplies were draining much faster than expected under 'real race' conditions. They consequently had to be replaced during the race.

Fig. 2. The component parts of the first version of the BPS box

In this prototype version, only SMS (Short Message System) messages were used to transfer the position data from to the service domain. Although GPRS (General Packet Radio Service) data transfer in principle was available, we chose not to use that. Firstly because the race route is on the border with, and partly crossing into Germany, and therefore the cellular network along the route did at the time not have full un-interrupted GPRS coverage. Secondly, for GPRS we would have had to implement the PPP, IP and UDP protocols in the microcontroller. This is certainly possible, but not within the short time vailable before the 2003 race.

Despite it advantages, there are also some major draw-backs to SMS. Firstly, there is a limited message size and character set (160 7-bits characters) and secondly the price per data unit is very high. The latter was of less concern, since the telecom provider Telfort (http://www.telfort.nl) was sponsoring the pilot by covering these costs.

In figure 3, a more detailed view of the architecture of the *service domain* is depicted. The *position database* is the central part of the service domain. In the pilot, we used the popular open source database server MySQL [4]. Unfortunately, it transpired that we required the use of sub-queries, which were only supported in the then latest version, MySQL 4.1.0 alpha. Using an alpha version for anything other then testing is of course never advisable, and it would turn out to be a bad choice indeed. During the race whenever database usage became more intensive, and as soon as the number of position entries became more than a couple of thousand, the database server locked up completely. The only solution available was cleaning up the table contents at regular intervals.

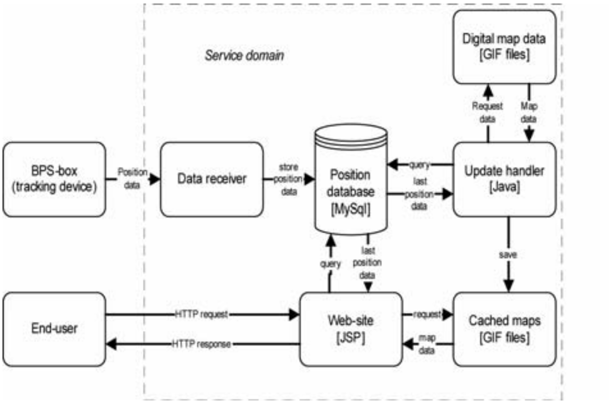


Fig. 3. Service domain architecture of BPS 1.0

The goal of BPS is to give anyone interested an insight in the progress of the race and for that just the raw position information is not nearly enough: No-one will be helped much by lists of latitude/longitudes. Therefore we needed to project these positions onto a map. Figure 4 shows a screen shot of the interface with the zoomed map. The team selected by the user is centered in the middle of the map, represented by an arrow. The angle of this arrow is defined by the travel direction retrieved from the position database. Any other team whose position is within the viewed area is shown as a red dot. Using the input list in the bottom of the screen, a user can choose to center the view on another team, or go to an overview map.

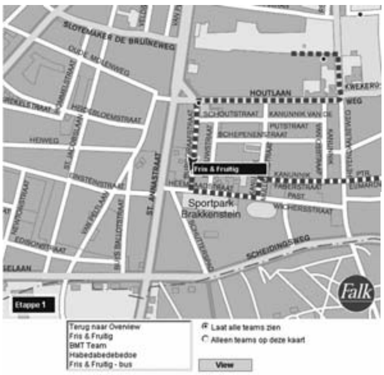


Fig. 4. User interface of the BPS 1.0 pilot

The maps used for displaying the position of the runners where made available by another sponsor (Wegener Falk, now Falkplan-Andes BV, http://www.falk.nl/). These were raster files of their commercial street maps, tiled into convenient chunks and stored in GIF format on the server. From these base files, the *update handler* constantly creates a cache of map fragments centered around the team's positions. This caching improves the performance of the *web site* JSP component (Java Server Pages), making the system more scalable. When an end-user requests the current positions on the website, the JSP will retrieve the latest positions and other relevant information from the position database and use that to determine which cached map data to retrieve and how to show it in the user interface for the *end-user*.

Despite the teething troubles mentioned, the general setup of the system, tracking runners and offering the collected information on the internet in real time, worked satisfactory. It was certainly promising enough to further develop a more reliable working system.

5 BPS 2.0 – the current system

The BPS 2.0 system as it is available at the time of writing is considered a stable version. It has been tested at the April 2004 Batavierenrace and further refined during subsequent events, taking place in different parts of The Netherlands. Based on the experiences with the protoype, it was decided to leave the general setup of the system unchanged, and concentrate on making it more robust and user-friendly.

For the *track domain*, a much smaller BPS-box has been developed (figure 5). With a size of only 12 x 6.5 x 3.6 cm and a weight of 200 g, it can comfortably be carried by a runner. When the internal battery is fully loaded, it will function for approximately 8 hrs. Additionally, external battery packs or other sources of energy (such as a 12 V car outlet) can make for an almost endless battery life. Instead of the standard GSM phone there's now a small on-board component and the hand-built microcontroller was traded in for a commercially available one with embedded Linux (eCos, an open source, real time operating system intended for embedded applications). This allows for flexible, fast and updateable controller logic.

Fig. 5. BPS 2.0 tracking boxes (12 x 6.5 x 3.6 cm). On the one in front the connection for the battery loader or the optional external power source can be seen. On top of the other box, the GPS antenna is visible.

GPRS is used, instead of SMS, as the primary means of data transfer, making more frequent updates feasible, as the price per data unit is much lower. Also there are no restrictions on the data representation: one can use any message format, with any character encoding. In case GPRS connectivity fails, the system can fall back on SMS until GPRS coverage is restored.

The position database now is implemented on the PostgreSQL database server [5], even though MySQL now does support sub-queries in its stable versions. The main reason for the switch was to provide for the future deployment of a spatially enabled database (see next section).

The client-side of the system was enhanced also. The website now can offer multiple zoom levels and provides the end-user with more information, during as well as after the race, such as replays of the tracking and speed diagrams (as shown in figure 6). Furthermore, at the highest zoom level not only maps are provided, but also aerial photographs, courtesy of the sponsor Aerodata International Surveys (http://aerodata-surveys.com/).

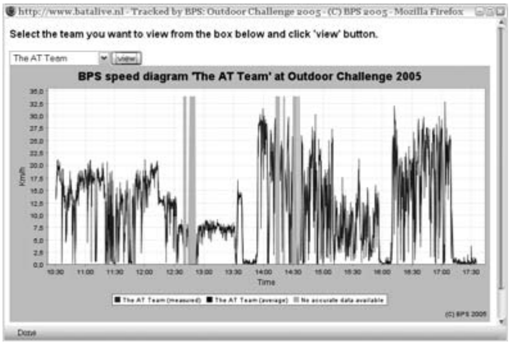


Fig. 6. Example of a speed diagram, generated during the Outdoor Challenge 2005, an orientation race with a combination of running, canoeing, cycling and kick-biking.

The BPS 2.0 system provides the real time position of the runners during the race and displays their positions on a map. As such, it is described most accurately as a Location Service (LS). Using the type of data we accumulate, it is also possible to combine that LS with all kinds of other spatial data and services, giving additional feedback to the runners, their teams and the spectators. As an example, the system could warn the runners when they are off-track, or let them know how close the nearest competitor is, or how to get to the nearest hospital in case of an accident. Offering this type of services, the system would turn into a genuine Location Based Service (LBS).

6 BPS 3.0 prototype – Development of a real time LBS

There are many definitions of an LBS around, a common one being LBSs are information services accessible with mobile devices through the mobile network and utilizing the ability to make use of the location of the mobile device [6]. The BPS system is theoretically able to provide the Location Based Services mentioned above. In what manner this can best be implemented was investigated as part of the research for the Bachelor Thesis in [3], and based on this a prototype has been developed. This prototype uses the same basic setup as the BPS 2.0 system, but it differs considerably in two aspects:

1. the geographical data is stored not in (raster) *map* form, but as geographic vector *data* in a spatially enabled database, and

2. this data is depicted in a user interface based on the Scalable Vector Graphics (SVG) format instead of raster GIF images.

The prototype is built using an n-tiered distributed design, with the possibility to spread tiers across servers, in order to improve scalability and performance, as shown in figure 7.

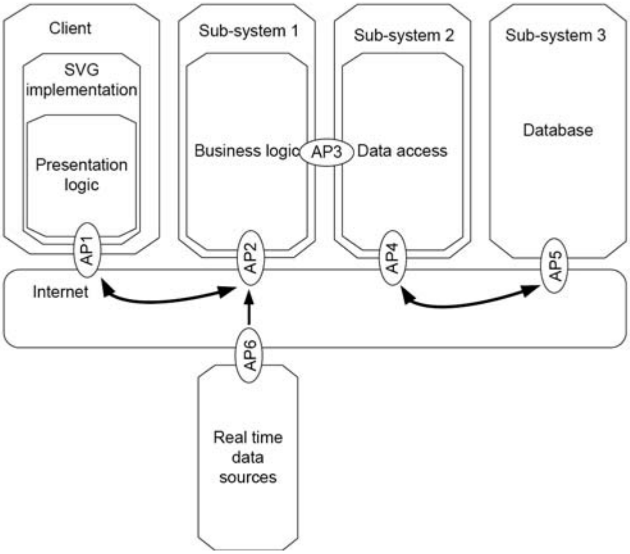


Fig. 7. The tiered architecure of the BPS 3.0 prototype (AP = Access Protocol).

The *Database Tier* is where the actual data storage is performed. The current BPS 2.0 stores in its database only data of the runners' locations. The topographic information is in the GIF map files but only implicitly, for viewing purposes, as the system itself has no knowledge of the spatial components of these maps (other than the georeferencing). One obvious problem with the current non-spatial database is that the system is fixed on data and maps projected in the Dutch national grid (Rijksdriehoeksstelsel). A conversion of the system to use non-Dutch maps would be possible, but it would have to be redone for every new projection. More importantly, the current system lacks support for *spatial queries*. Because of that, it is impossible to ask questions like "find all runners within a certain area", or "how far ahead am I of my nearest competitor?" to the database.

If one uses a database to serve as the backend of a geospatial service such as BPS, the most obvious choice is a so-called *spatially enabled* database, that is one that has special data types, functions and operators to deal with real-world geometries. Such spatial extensions, or spatial cartridges as they are sometimes called, provide access to the geometry through an SQL interface. Open standards for this interface have been set by the Open Geospatial Consortium (OGC,[7]). There are several databases with OGC-compliant spatial extensions. PostgreSQL (when using the PostGIS plug-in [8]) and MySQL are open source examples thereof. PostgreSQL/PostGIS offers a more mature and complete implementation, and furthermore MySQL spatial functionality is only implemented for the MyISAM table type. MyISAM does not support more advanced database features such as foreign keys, triggers and views, and especially lacks *transaction* capabilities, which are essential in a multi-user system such as BPS.

The database serves multiple roles in the system: Firstly, it stores and retrieves spatial data of the topography, which is then used to generate SVG maps. PostgreSQL/PostGIS enables generation (and caching) of the chosen output format SVG (see below) natively in the database. This saves a lot of conversion time later on, when hundreds of requests are made for the same piece of data.

Secondly, it contains functionality related to the processing and storage of the real time positioning data. By deploying custom-written internal functions, we have been able to cut down the number of separate queries necessary for complicated actions like setting up a new participant. This greatly reduces the time needed for client–server communication.

The system accesses the database through the *Data Access Tier*. This interface to the data tier handles all data input and output. Its main task is to keep track of all data sources and provide access to these through a Connection Pool. The mechanism permanently keeps a number of connections to the database open, ready for immediate use and avoiding the continuous and time consuming opening and closing of connections. We've also implemented a form of priority queuing, as some data requests are more important than others, and are thus given precedence when resources are scarce.

The *Business Logic Tier* handles incoming requests, by figuring out exactly what data is needed to answer the request, making sure the requested data is received, and transforming it into a valid response.

The communication between the tiers is realised using *Acces Protocols* (AP*n* in figure 7). These follow the Web Services paradigm, being regular HTTP requests, using a pre-defined set of GET variables. Requests are received by one servlet that is a part of our business logic layer (with a URL in a format of http://serveraddress/Request?request=TheRequest). This servlet then distributes the request to the class responsible for answering that specific type of request. The available requests are

RealTime, GetMap, GetParticipants and *GetNearest*. In these interfaces, we implement selected features from OGC's Web Map Services specification (WMS, see [9]) and use its nomenclature and, if possible, semantics for parameter names and behaviour where applicable. Normally, all requests to the server are made by client tier, but it is possible to use these interfaces by calling them directly from any application (eg. any web browser).

A special case among these interfaces is the *RealTime* interface, since it does not request any data from the server, but rather submits new data of the runners' position. The original plan was to try to comply with the OGC Open Location Services specification [10]. However, on closer inspection we noticed that the fundamental setup of this standard is that a client is unaware of its location, and asks the server to determine its position. This is the reverse of how BPS works, where the clients (in this case the tracking boxes) know exactly where they are, and want to transmit this information to the server. Therefore we decided to use WMS style parameters for this interface instead.

The *GetMap* request is used to retrieve map data in vector format from a database. Which databases are available, and how and from which tables this data is gathered is defined in a configuration file. The GetParticipants interface generates a listing of the latest available data for participants that have a database entry. This includes data such as a participant’s position, name, stage and speed. The GetNearest interface can be used to locate the nearest hospital, first aid point, competitor or other object relative to a given participant. It returns the Euclidian distance to these objects.

The *Presentation Tier* is physically on the client’s machine. This component is the end user view of the system and includes the Graphical User Interface (GUI). With the original BPS system, we did put almost no requirements on the client browser and therefore the obvious choice was to use raster GIF images. This enables all users to see the maps without the need for additional plug-ins. But using raster means maps are always transported in full over the Internet, and requires the browser to refresh all the map data regularly, even if there is no change in the runner’s location and no new map data is required. For BPS 3.0 we chose to use Scalable Vector Graphics (SVG, see [11]), for both the maps and the GUI itself. Using vector on the client allows the creation of a better-looking and more responsive GUI. Examples of this are the use of anti-aliasing on the client, and the loading of new map data in the background without the need to redraw the whole user interface.

In the end the vector data has to be rendered to the screen of course, putting some additional requirements on the client. In most cases, some sort of SVG plug-in is needed, although browsers with native SVG rendering are starting to appear (eg. FireFox 1.5 and Opera 9).

For the client architecture, we used the basic functionality of an existing application framework from CARTO:NET [12] and modified it to fulfill our needs. The part we had to add was the dynamical loading of the map data and the participant information. To have the client issue the requests to the bussiness logic, we've used the getURL() function that is currently only implemented in the Adobe SVG plugin, but is also included in the upcoming SVG 1.2 recommendation. The output of all three *Get* requests mentioned above are snippets of SVG vector data, which are parsed into the client SVG.

The end-user has the possibility to select a participant and zoom in on his location. One can then can choose to track the participant automatically or to see where the nearest competitor, hospital, ectetera is located. The participant's information is refreshed at a chosen interval and his position updated on the map. When the participant goes out of the current map frame, new map data will be retrieved and the map is panned accordingly.

In order to evaluate the merits and drawbacks of the prototype, we compared it to the current BPS 2.0 system. With regard to the *performance*, the first load of the prototype is slightly slower, since multiple requests need to be made to the system, but after this it is significantly faster at handling user requests. In terms of *scalability*, the prototype is comparable to the current BPS. Both can be spread over a large number of servers, using Tomcat's built-in load-balancing features and replication for the database. The *extensibility* of the prototype is better, due to a much better separation between presentation and business logic, and the way in which requests are structured. The prototype's new *functionalities* will give the end-user more interaction and enables all kinds of spatial queries which are not possible with BPS 2.0. The use of SVG in the client makes the system less platform-independent on the one hand, but generally the use of vector data in both server and client gives a lot of new possibilities, which we haven't explored fully yet.

7 Conclusion

The BPS system combines off-the-shelf components and combines them with open standards compliant internet applications into a highly efficient and powerful application. The current BPS 2.0 is a low cost and scalable system, making it possible to track many objects simultaneously and display their positions on the Internet in real time. It has proven its useability in several use cases under real-world conditions.

In its current BPS 2.0 version, it can be argued that it principally provides Location Services, but it has a great potential for Location Based Service (LBS). The prototype BPS 3.0 that we presented still needs some fine-tuning and can be improved in several ways, but already we think its proved a solid basis for further development of a flexible, multi-purpose, genuine real time LBS.

References

[1] Batavierenrace website. http://www.batavierenrace.nl/english/ (last accessed: February 2006).
[2] Ekkebus, S. P. (2003): Realtime location based positioning system - a requirement study and design. Bachelor thesis. Enschede: University of Twente, EWI. 63 p.
[3] Boekestein, D. & P. G. Uithol (2005): Location based services using database driven vector based graphics. Bachelor thesis. Enschede: University of Twente, EWI. 72 p.
[4] MySQL AB website. http://www.mysql.com/ (last accessed: February 2006).
[5] PostgreSQL website. http://www.postgresql.org/ (last accessed: February 2006).
[6] Virrantaus, K., J. Markkula, A. Garmash & Y.V.Terziyan (2001): Developing GIS-Supported Location-Based Services. Proceedings of WGIS 2001 – First International Workshop on Web and Wireless Geographical Information Systems. Kyoto. pp. 423–432.
[7] Open Geospatial Consortium: OGC Home Page. http://www.opengeospatial.org/ (last accessed: February 2006).
[8] Refractions PostGIS website. http://postgis.refractions.net (last accessed: February 2006).
[9] Open Geospatial Consortium Inc. (2002): Web Map Service Implementation Specification (1.1.1) OGC 01-068r3, J.d.L. Beaujardière (ed.).
[10] Open Geospatial Consortium Inc. (2004): OpenGIS Location Services (OpenLS): Core Services (1.0) OGC 03-006r3, M. Mabrouk (ed.).
[11] World Wide Web Consortium (W3C): Scalable Vector Graphics (SVG) 1.1 Specification. http://www.w3.org/TR/SVG11/ (last accessed: December 2005).
[12] carto:net site. http://www.carto.net/ (last accessed: March 2006).

Smartphone-based information and navigation aids for public transport travellers

Karl Rehrl[1], Nicolas Göll[1], Sven Leitinger[1], Stefan Bruntsch[2], Hans-Jörg Mentz[3]

[1]Salzburg Research Forschungsgesellschaft mbH, 5020 Salzburg, Austria
{krehrl, ngoell, sleiting}@salzburgresearch.at
[2] Verkehrsverbund Ost-Region GmbH, 1060 Wien, Austria
stefan.bruntsch@vor.at
[3]Mentz Datenverarbeitung GmbH, 81675 Munich, Germany
mentzh@mentzdv.de

Abstract

Digital navigation assistance on Smartphones has recently gained high attention due to capable mobile devices, off-board based navigation software and advancements in wireless network technologies. Mobile information and navigation aids can assist travellers in non-trivial navigation and orientation tasks in unknown geographical regions or transport systems. Whereas car navigation has reached a certain level of maturity, assistance for multimodal travellers especially in public transport networks is still in its infancies. Integration of mobile multimodal journey planning and navigation and guidance in complex public interchange facilities is not adequately addressed by existing systems. Thus, in this paper we propose a personal travel companion on Smartphones for assisting multimodal travellers. We focus on the topics mobile journey planning, pedestrian route calculation, navigation and guidance in public interchange buildings and seamless transition between indoor and outdoor positioning. Moreover the paper describes the prototypical implementation on Off-the-shelf Smartphones and evaluation results.

1 Introduction

One of the most important needs of society today and tomorrow is mobility. Mobility is always closely related to travelling, describing a person's task of moving from place A to place B. Travelling can be done using different transport means, including individual motorised transport or public transport. Travelling to foreign places is considered as non-trivial navigation task [26]. Wayfinding in unknown geographical regions demands spatial knowledge which can be provided by navigation aids [5],[23]. Recent technological advancements in the field of Location-based Services and Telecartography build the basis for digital navigation assistance.

While more and more car drivers are used to relatively mature guidance and navigation systems, providing trip information, guidance and orientation for public transport travellers is still an open issue. Investigations have shown that public transport systems are complex [16] and often provide obstacles for travellers which are not used to public transport systems. Digital information and navigation aids can provide continuous on-trip assistance and thus overcome existing barriers.

When we take a closer look at travelling in the public transport system we can distinguish between two typical travel situations: travelling with public transport means where travel routes are fixed in time and space and wayfinding situations in interchange facilities or in the surrounding areas of public transport stops. Digital information and navigation aids for public transport systems have to cope with both travel situations. Providing overall trip information containing information on timetables, transport means, interchanges and final stops requires trip management on mobile devices. Finding the way from an address to the nearest transport stop or from the final transport stop to an address requires outdoor pedestrian navigation. Finding the way in public transport interchange facilities requires indoor navigation. A combination of both is needed if the pedestrian route starts in an interchange building (e.g. an underground station) and leads to an address or public transport stop over ground.

Due to the constantly increasing technical advantages of Smartphones, pedestrian navigation recently has gained high interest as one of the potential mobile killer-applications in the near future. However, pedestrian navigation on Smartphones is still in its infancies. Most of the commercially available systems were originally designed as car navigation systems and are now sold as pedestrian navigation systems with only minor modifications (e.g. Wayfinder Navigator [36]). An integration of public transport trip management and indoor as well as outdoor guidance is not yet achieved by existing systems. The missing link is a mobile traveller assistance system for palm use providing on-trip information and guidance according to specific travel situations on multimodal journeys[1].

Having this vision in mind, a consortium of Austrian and German transport associations, research organisations and companies, led by the biggest Austrian public transport association, the Verkehrsverbund Ost-Region (VOR)[2] has worked over the last three years to integrate and extend existing travel information systems in order to achieve continuous personal on-trip assistant for public transport travellers. The goal was to overcome the deficiencies of heterogeneous, not integrated systems for multimodal journey planning and pedestrian navigation in order to provide new approaches to passenger guidance and information, which should increase the attractiveness of public transport.

Our approach to a multimodal travel assistance application on Smartphones combines two modules. The first module is a browser-based mobile access to a server-based multimodal journey planner which allows users to calculate multimodal routes between given start and end points. The result is composed of individual trip segments tagged with information about type of transport and estimated travel time. The second module provides an off-board navigation service that guides the user on outdoor as well as indoor pedestrian routes. Both components are integrated in a mobile application called the personal travel companion that can be accessed by public transport travellers whenever and wherever they want or need to.

The main questions answered in this paper are how to bring multimodal journey planning and trip management to mobile devices, how to enable location-based journey planning on the Smartphone, how to provide guidance and orientation to passengers in interchange situations and how to assist travellers in finding their ways to and from addresses or public transport stops.

The paper is structured as follows: First we describe requirements for providing information and navigation aids for public transport travellers. Building on these requirements we describe related work as well as the design and prototypical implementation of a personal travel companion for Smartphones. We focus on the topics mobile journey planning, route calculation, navigation and guidance in and in the surrounding of interchange buildings and positioning. We finish with concluding thoughts.

---

[1] Using different transport means on a single journey
[2] http://www.vor.at/

2 Requirements

Examining typical travel situations of public transport passengers we have identified four requirements which are crucial to the design of a personal travel companion:

- Mobile trip planning and management
- Detailed modelling of pedestrian networks
- Navigation and guidance for public transport travellers
- Continuous positioning

These requirements were considered as the key research questions.

2.1 Mobile trip planning and management

Regarding this requirement we mainly focus on extensions of existing multimodal journey planners concerning mobile trip management and planning. Mobile trip management means that a personal travel companion on a mobile device allows to access personal trips planned in the pre-trip phase, e.g. by the use of a web-based multimodal journey planner[25], to store personal trips locally on the mobile device or on a server and to keep a personal history of planned routes or preferred travel locations. Mobile trip planning includes location-aware journey planning via mobile access to a server-based journey planner. Results from the journey planner should provide optimised route suggestions from the current location to a destination address under consideration of connections in the public transport network and pedestrian routes for interchanges. If available, personal travel preferences as well as real-time data from operational control systems should be considered.

2.2 Detailed modelling of pedestrian networks

Multimodal web-based journey planners have recently come to maturity and thus gained considerable attention [33],[31]. Whereas typical multimodal routes consist of at least car, public transport and pedestrian parts, detailed data of pedestrian networks is often poor or missing. Especially when it comes to complex interchange buildings, interchanges are calculated on the basis of an interchange matrix, connecting different transport means logically. However, in order to provide navigation and orientation aids in these situations, detailed modelling of interchange facilities and pedestrian routes is necessary.

Concerning the surroundings of interchange facilities, a fine-grained geographical modelling of pedestrian routes and the surrounding environment is necessary. The typical scales used for modelling the road network for car navigation are to coarse-grained in order to cope with the requirements of pedestrian navigation [6].

2.3 Navigation and guidance for public transport travellers

Increased complexity [16], information overload and unknown situations often make people feel uncomfortable while travelling with public transport. Especially when public transport means are left for interchanges people are confronted with complex wayfinding tasks. A key requirement for a personal travel companion is to provide orientation and guidance in complex interchange situations. Experiences with car navigation systems have shown that navigation tasks can be successfully assisted by electronic navigation devices. Until now navigation systems for public transport passengers are missing, in spite the fact that orientation with unknown public transport systems is not easier as it is with road networks. From a technical point of view, providing navigation and guidance in complex interchange facilities as well as from public transport stops to addresses is a challenging task. Pedestrian navigation in combined indoor/outdoor environments is still an open question of research. The key questions address positioning, description of buildings, transition from indoor to outdoor and vice versa and the generation of route descriptions considering the surrounding environment. Integrating the combined indoor/outdoor pedestrian navigation into the mobile journey planner on Smartphones was one of the most challenging tasks during the development phase of the personal travel companion.

2.4 Continuous positioning

As mentioned in the section above, the means of obtaining the user's position is a key question of every location-based services. The task of positioning is also particularly difficult for public transport travellers. Looking at typical interchange buildings in urban areas we face the problem that most of the pedestrian routes are indoors. Although considerable advancements to the use of GPS in buildings have been made [8],[32], GPS technology is still not capable of replacing indoor positioning technologies. Finding suitable indoor positioning technologies for public interchange buildings, allowing a seamless transition from indoor to outdoor and vice versa was considered a main requirement for our approach to a personal travel companion.

3 Related Work

According to the key research questions described in the requirements section we classify related work into the following categories: multimodal journey planning, theory of wayfinding in public transport buildings, pedestrian navigation pilot systems as well as outdoor and indoor positioning technologies.

Multimodal journey planning is a relatively well researched area and some of the research prototypes are already in commercial stage. EU-Spirit [10] is a European travel information system offering calculation of door to door travel itineraries between European cities or regional areas. Besides the operational system results include open interface definitions and harmonised metadata. The main goal of the ISCOM [18] project was to model the ways to and from public transport stops in a geographic reference system in order to integrate these paths in the multimodal routing network. Results from the ISCOM project are the basis for the intermodal journey planner (IJP) provided by Mentz Datenverarbeitung GmbH[3] which was used for our prototype applications. However, none of the projects focuses on the combination of mobile multimodal trip management and pedestrian guidance along public transport routes.

The human navigation and wayfinding process is based on concepts of human cognition [7],[15]. Rüetschi and Timpf [29],[30] developed a conceptual model for describing the wayfinding process in public transport stations. They differentiate between the network space (the public transport network itself) and the scene space (the nodes of the public transport network, e.g. interchange facilities). The scene space is modelled by the schematic geometry, which is based on image schemes [19] and affordances [13]. In another study Fontaine und Denis [11] analyse spatial human cognition in subway stations. One of the results of the study with several users is that direction signs are particularly important elements for the navigation and wayfinding in public transport stations. Signposts are significant elements for the orientation at decision points. This result is also confirmed by May et al. [23] in a requirement study of pedestrian navigation. Our approach builds on theoretical concepts of wayfinding, focusing on the design and implementation of navigation aids. Especially the incorporation of path segment types and signs in the generation of

[3] http://www.mentzdv.de/

maps and guidance instructions addresses the peculiarity of human spatial cognition.

There are many pilot projects which tackle different topics of pedestrian navigation, like user interaction, positioning, cartographic visualisation and data transfer. REAL [1] and M3I [20] are developed for indoor and outdoor environments. REAL describes a hybrid navigation system that adapts the presentation of route directions to different output devices and modalities. M3I presents an approach that connects a variety of specialised user interfaces to achieve a personal navigation service spanning different situations. The NAVIO project [13] analyses major aspects being important when designing a pedestrian navigation system for indoor and outdoor environments. The main parts of the project are integrated positioning technologies, multi-criteria route planning and multimedia route communication. LoL@ [4] is a pedestrian navigation system for tourists which operates on Smartphones. LoL@ concentrates on cartographic visualisation of multimedia content. As far as we know, there is no existing pilot system which focuses on mobile multimodal route planning in combination with navigation and guidance of public transport passengers in interchange facilities.

Addressing the field of positioning Retscher and Thienelt [28] discuss suitable location technologies for pedestrians. In their study they test and demonstrate different positioning technologies like satellite-positioning, cellular phone positioning, dead reckoning sensors for measurement of heading and travelled distance as well as barometric pressure sensors for height determination. For indoor positioning most of the prototypes or commercial systems are based on Infrared, WLAN or Bluetooth [11], [17], [20], [35]. Whereas Infrared needs line of sight, WLAN positioning needs costly calibration and can not be accessed by typical Smartphones. Bluetooth positioning systems are mainly server-based and thus require a costly installation procedure. In our approach the main requirements were defined by an easy and cheap installation procedure and the use of Off-the-Shelf Smartphones.

4 Design of the personal travel companion

Starting from the requirements outlined above the design of a personal travel companion for mobile use focuses on detailed pedestrian route calculation as extension to coarse-grained multimodal route calculation, navigation and orientation in complex interchange buildings as well as positioning techniques.

4.1 Pedestrian route calculation and modelling of interchange facilities

Multimodal route calculation is done between two points. These points may be addresses, entrances of a POI (point of interest) or public transport stops. The route is calculated on integrated public transport, road and pedestrian networks. The networks are connected at defined transition points and are based on a common geographic reference system.

An important topic of research within the project was the very detailed calculation of pedestrian routes in public transport interchange buildings. A pre-requisite for this task is a model for interchange buildings including a routing-enabled pedestrian network.

For modelling interchange buildings we have adapted and extended a conceptual model originally proposed by Ruetschi and Timpf [29], [30]. The model differentiates between the network space (the public transport network itself) and the scene space (the nodes of the public transport network, e.g. interchange facilities). The scene space describes the space where wayfinding in public transport interchange facilities takes place. The model is based on a hierarchical, logical representation of the real world which is capable of modelling the complex spatial relationships of structural elements in interchange buildings.

Our adapted model uses the following main data categories: building, floors, regions, gateways and items. This logical representation of buildings is necessary in order to model the pedestrian routing network in detail, position pedestrians along interchange routes, show floor plans and give meaningful routing instructions. Interchange buildings are divided into different floors with a hierarchical order. Floors are logically structured in non-overlapping walkable regions. For the connection of floors and regions gateways are used. Typical representations of gateways are stairs, elevators, escalators or ramps. For better orientation it can be useful to collect data of certain items. These may be signs, ticket machines, shops etc. Items can either afford a user interaction like ticket machines, or they can be used as orientation marks like signs. They are linked to regions and are used to provide a better interaction between wayfinders and the surrounding environment.

The logical model of buildings is augmented with a geographical model based on a coordinate system. The resulting model is called a hybrid location model [21]. All regions are modelled as non-overlapping polygons, so-called zones. Gateways have gateway areas (polygons) in the origin region and target coordinates in the destination region. The coordinate system was extended with a third parameter called level, which indicates the floor of the building.

In order to enable the calculation of pedestrian routes, a pedestrian routing network is added to the model. This graph-based network consists of nodes and segments and has to be linked to the logical model of buildings and to road and public transport networks. To each node or segment we are able to assign certain attributes which are used for route computation and generation of routing instructions. Nodes are marked with a floor number. When a path is computed, the order of floor numbers provides details whether the segment between two nodes leads up or down.

Segments have type attributes, which indicate the type of a segment like an ordinary path (even level), stairs, escalators, elevators or ramps. With this information we are able to realise a selective route computation based on personal demands in order to optimise interchange time, route complexity or walking effort [16]. Additionally, each segment has to be tagged with a direction because of the possibility that escalators are only available in one direction or the direction can be changed. For wheel chair users it is important to acquire details in the building structure like single steps or the grade of ramps as well.

Another important attribute is the time needed to walk along a certain segment. The overall interchange time has to be considered for continuous route calculation. In order to provide personalised interchange times we use time factors for each path segment. During the route calculation these time factors are multiplied by the default velocity settings in the traveller’s personal profile.

In order to give travellers detailed route instructions for orientation inside the building, data about existing signs are collected and linked to the directed path segments. After computation of an interchange route relevant signs are selected and the corresponding text of the sign is included in the route description. Selection of relevant signs from all available signs is only possible if they are linked to exits and public transport transfer points of the building. This technique lays the ground for personalised guidance as an interaction between digital and physical guidance systems. Moreover, the model is the foundation for automatic positioning along the route and indoor/outdoor transitions.

4.2 Navigation and orientation in complex public transport interchange buildings

Starting from the conceptual model of buildings in this section we describe our approach to navigation and guidance in interchange buildings. We define guidance as an information technology based tool assisting pedestrians in the process of wayfinding, which means a purposeful interaction with an environment where the purpose is to reach a certain place or goal [29], [30]. Our guidance system provides the following services:

- select the optimised pedestrian route according to the user's profile in interchange buildings (e.g. from the car to the platform or from one platform to another)
- give instructions for pedestrians in order to optimise their interchange and to improve their orientation
- select the most relevant information out of the scene space based on the calculated footpath in order to improve the interaction between wayfinders and the environment [12]
- reduce complexity of the pedestrian's navigation task by providing detailed timetable information of public transport connections, detailed floor plans of interchange buildings, automatic positioning and thus filtering of information from an egocentric view [24]

The process of guidance in interchange buildings can be described as follows: Upon request travellers in interchange situations get a detailed description of the interchange building, the logical model of the building, floor plans, the calculated interchange route, route instructions and positioning information. After the user's current position has been determined, the location within the model is calculated and the according floor plan together with the relevant route instructions are displayed (Fig. 1). Floor plans are simplified representations of the scene space, including floor numbers, walkable regions, calculated route segments, gateways, signs and optional orientation marks. Beside general information on the building, they typically only show data relevant for the calculated path. Thus floor plans have to be generated dynamically from the model.

Regarding the route instructions it was important for us to avoid simple turn-by-turn instructions that are solely based on geometric information that result in guidance texts like "Walk nine meters straight and turn left." Instead of that instructions should refer to observable physical objects in the scene space in order to improve the interaction of pedestrians and the environment [12]. Referenced objects can be gateways, signs or orientation marks. The generation of route instructions is based on a set of standardised text building blocks which allow us to create appropriate path descriptions for most cases. For complex scenes it is possible to link manual route directions to specific path segments. This basic path description is combined with information from nearby landmarks and signs that are stored in the database. In doing so it is possible to reference signs that do not explicitly refer to the traveller's destination but point at the right direction. In this fashion we are able to automatically generate instructions like “Walk to the lower end of the stairs marked with the sign ‘Neubaugasse’. Walk up the stairs.”

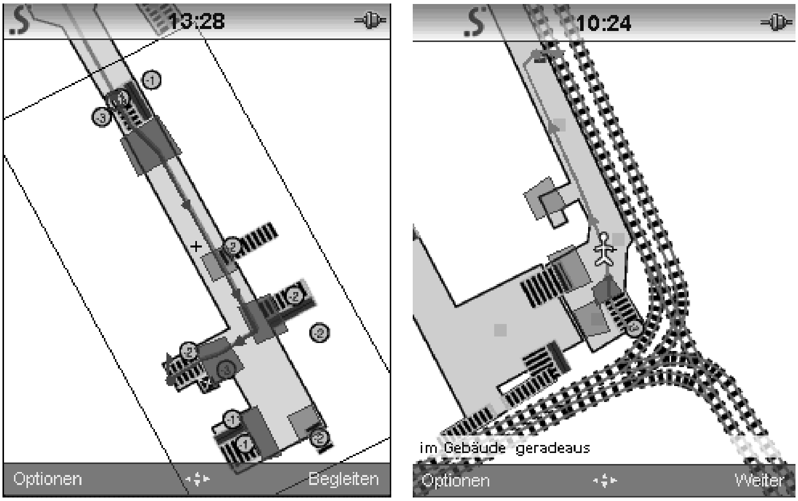


Fig. 1. Screenshots showing the map-based guidance in interchange buildings (manual and automatic navigation mode)

In addition to indoor guidance we address outdoor guidance as well. Outdoor pedestrian guidance is necessary for interchange facilities having bus or tram stations over ground or for finding the way to the destination address from the final public transport stop. For pedestrian routing and navigation the road network as well as separately modelled pedestrian networks can be used. Pedestrian networks can either be extensions to road networks (e.g. sidewalks, pedestrian crossings or pedestrian underpasses) or road independent networks (e.g. interchange buildings, parks, or pedestrian zones). The important difference between road and pedestrian networks is, that the routing network and the surrounding environment have to be modelled with different granularities. Whereas for car navigation it is sufficient to model roads in large scale (e.g. coarse-grained road segments with attributes), for pedestrian navigation paths and the surrounding environment have to be modelled in detail in order to be of use for navigation and orientation purposes [6]. Because of the slow movement of pedestrians the perception of the environment is significantly increased and thus local landmarks and detailed walkable regions in scales of 1:500 and below should be shown on maps and referenced in route instructions. E.g. modelling of over ground interchange facilities has to include public transport platforms, public transport infrastructure like tracks or bus stops, walkable pedestrian regions, non-walkable regions, prominent buildings or other objects, connections to sidewalks of nearby streets and pedestrian crossings. Maps or instructions generated on the basis of road network data can not meet these requirements, thus it was a challenging task to adapt the models and to generate fine-grained maps and routing instructions. Whereas a granularity of one meter is sufficient for car navigation systems, a pedestrian navigation aid for indoor use sometimes requires the model to be precise up to the range of centimeters, especially when it comes to the generation of fine-grained digital floor plans.

4.3 Positioning

Positioning is an integral part of navigation systems in outdoor as well as indoor environments. Although our guidance system provides manually navigable maps and a list of step-by-step route instructions which can be manually acknowledged, the maximum convenience for travellers can only be achieved with automatic positioning. In order to cope with the requirements for continuous positioning we had to find solutions to positioning for indoor as well as outdoor use. Whereas outdoor positioning in variable, yet sufficient quality is available by the use of global positioning systems (e.g. GPS), indoor positioning has not yet reached the same level of maturity. For indoor positioning numerous different approaches that vary greatly in terms of accuracy, cost and used technology [17],[23] exist. In order to be applicable for our scenario we determined the following criteria:

- to provide high enough accuracy to determine the region where the user is currently in
- to have broad support of end user devices
- to work without cellular network connection
- to be cost effective
  to require little installation effort

We opted for a Bluetooth based solution because it met our requirements most closely. First of all a great share of Smartphones sold today incorporates this technology and thus will support automatic positioning without additional hardware on the client side. Furthermore we felt confident to reach a high enough accuracy for providing orientation and useful instructions for the wayfinding process.

Most of the commercially available location systems based on Bluetooth (e.g. [2],[3],[22]) use an infrastructure of interconnected Bluetooth access points. These access points permanently execute inquiries in order to detect nearby Bluetooth devices. Once discovered, their location is determined on the server side and appropriate information is pushed onto the detected device. In large public transport stations, however, it would be very resource consuming or even impossible to install a LAN interconnecting the access points.

Our approach to a cell-based positioning system makes use of a client-side inquiry and a set of passive Bluetooth beacons. The Smartphone clients are constantly looking for beacons in the proximity that broadcast their unique ID. After receiving a beacon ID, the client looks up the associated position information in a list that is part of the building description. If all the relevant data for an interchange building is cached on the device, navigation will work without any network connection. This is a crucial requirement for underground stations suffering from low cellular network coverage.

Fig. 2: Bluelon BodyTag

The beacons utilised for the pilot system (BlueLon Bodytags [3], figure 2) have an adjustable transmit power which allows for adjusting cell sizes. This way we are able to adapt it to the needed accuracy or to the room topology at hand (i.e. hall, room or corridor). Ideally cell sizes should be selected in a way that the covered area is not overlapping with other beacons, otherwise this would result in an ambiguous position.

Due to the signals' spherical propagation behaviour it is not always possible to completely separate individual cells (i.e. signals crossing floor bounds). To overcome this problem we exploit the data model's hierarchical nature and use knowledge from the calculated path as well as information known from history. In a first step we sort out detected beacons that are outside the current region. Furthermore we can determine a sequence of beacons that will be passed when walking along the calculated route.

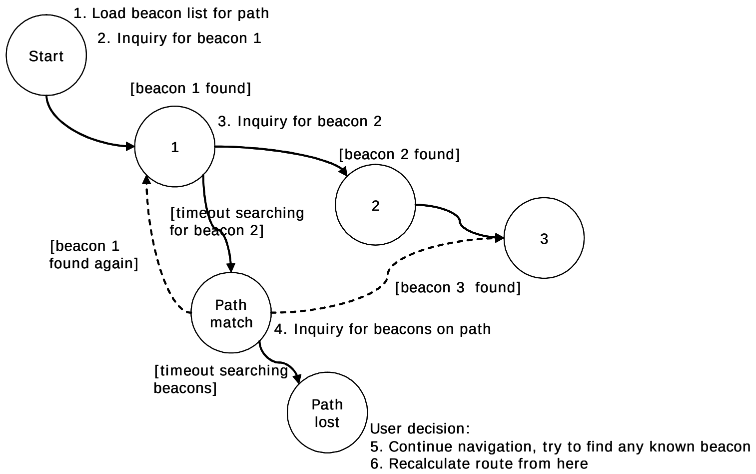


Fig. 3. State transitions of indoor navigation algorithm

If still more than one beacon is recognised and one of them is the next expected beacon, it is assumed that the user most probably moved one step further along the way. Likewise, if the next logical beacon is not found but the one following thereafter, we consider one beacon has been skipped. This procedure is illustrated in Figure 3.

Another challenging characteristic of the Bluetooth technology is the rather long delay from entering a device's transmit range until its actual detection. This can take up to 13 seconds and imposes a lower bound to the usable cell sizes, because a user may have passed the beacon without detecting it. However, tests have shown that most of the time beacons are found within the first five seconds. This observation is also confirmed by other studies [27]0. Experiments have further shown that restarting the inquiry after this duration yields higher detection probability. Together with the fault tolerance mechanism outlined above we achieved a usable cell size of down to 4 meters which is sufficient for providing useful route instructions. Improved inquiry performance is expected from Bluetooth 2.0. First tests have shown that inquiry time was reduced significantly and often successfully completed in one second or below. This would allow for smaller cell sizes and higher detection probability.

As described in the requirements section above, a crucial requirement for the personal travel companion is continuous positioning. Continuity in positioning means that transition from indoor to outdoor or vice versa is done automatically.

In our approach automatic transition is achieved through tagging regions with the positioning mode. Outdoor regions are mostly tagged with GPS whereas indoor regions are tagged with Bluetooth. This approach works well for outdoor to indoor transitions, where the navigation application simply disconnects from the GPS receiver and starts searching for Bluetooth beacons. Having found the first beacon the personal travel companion knows the indoor position and continues to give route instructions for the indoor route.

Transitions from indoor to outdoor are done in a similar way. When the last beacon before an exit is found the Smartphone client stops inquiry and connects to the GPS receiver. However, we are also facing the problem that currently available GPS receivers need an initialisation time of about 30 sec. up to several min. to deliver a reliable position. We try to address the problem by pretending to the user he is constantly moving along the calculated path until we are able to determine the exact location. In most of the cases this behaviour will lead to better results, because there is the chance of leaving the shadowed area during walk. We expect improved results with the broad availability of A-GPS [9].

5 Prototypical implementation

In this section we give some implementation specific details on the prototype application called the personal travel companion. The application is split up in server side and client side modules. For the multimodal trip planning on the server-side we use the Intermodal Journey Planner [25]. This service was adapted for calculating pedestrian routes and providing automatically generated pedestrian maps, path descriptions and data of interchange buildings. Communication between server and client is done via XML.

The Smartphone client was implemented with the mobile Java platform J2ME. The reference device is the Off-the-Shelf Smartphone Nokia N70. The architectural building blocks on the client are (Fig. 4):

- a micro browser for interaction with the multimodal journey planner and presenting server-calculated trip information
- a plugin mechanism for the micro browser to start specific extension modules (Smartlets and Services)
- a navigation plugin (Navigation-Smartlet) for handling all navigation specific user interactions
- a location service and two different location providers for automatic location acquisition
- a local data cache for caching trip and navigation relevant data

The main interaction concerning trip planning and browsing trip information on the client is done via a micro browser. This micro browser allows rendering of server generated pages similar to xHTML pages. However, advanced functionality on the client can not be handled appropriately by a standard xHTML browser. Thus, the browser is enabled for handling small functional extension modules called Smartlets. Smartlets are referenced in the markup language by proprietary tags. Upon activating a Smartlet link, the micro browser starts the corresponding Smartlet and hands over the control to this module. This mechanism guarantees a trade-off between server-generated user interfaces, where the design is easily exchangeable and additional functionality, which allows using local resources like positioning infrastructure on the Smartphone.

One of these Smartlets is the Navigation Smartlet. The Navigation Smartlet asks the local data cache for the journey data of the current trip including route descriptions, building descriptions and maps. Maps are delivered as geo-referenced bitmaps or vector graphics and can be split up in single tiles in order to improve loading times. The local data cache either uses locally stored data or fetches data from the server. This mechanism allows for preloading all the data for a whole journey or at least for one interchange building. Different strategies for preloading of data are possible.

The Navigation Smartlet performs the entire map rendering itself and communicates with locally available positioning providers like GPS and/or Bluetooth providers. Communication with Bluetooth is implemented using the Java Bluetooth API (JSR-82), which is available on recent Smartphones. Communication with the GPS receiver is done via the NMEA protocol over a serial Bluetooth connection.

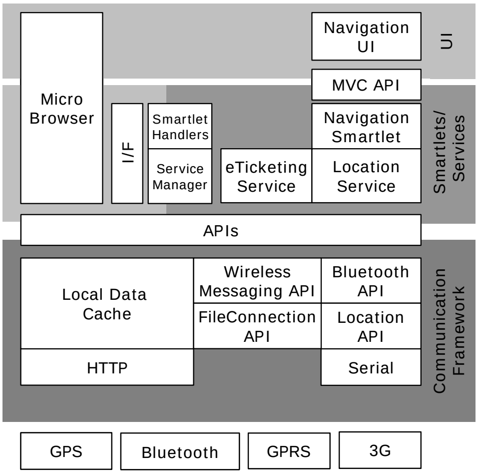


Fig. 4. Overview of the J2ME Client Architecture

A typical execution of the Navigation Smartlet is:

1. Determine the active journey part from the parameters passed via the Smartlet call
2. Load a list of instructions for the pedestrian route from the local data cache
3. Try to determine the current position automatically via the active location provider (Bluetooth or GPS)
4. Try to match the current position with the calculated footpath corridor
5. Get a map for the determined location from the local data cache
6. Render the map on the display and draw the current position on the map
7. Display the wayfinding instruction for the current position of the user
8. Repeat steps 3 to 8 until the user reaches the destination or exits navigation

Without automatic positioning the user has the possibility to scroll maps manually and to switch to other floors by interactively selecting gateways on the map. Moreover, users can also jump step-by-step through the instruction list and match the instructions with their actual position manually.

6 Conclusion

In this paper we described an approach to Smartphone-based information and navigation aids for public transport travellers. Surveys [34] have shown that disorientation on complex public transport routes is a typical barrier for travellers and thus will lead to a reduced use of multimodal transport. The assumption is that the share of multimodal transport can be significantly increased by providing digital mobile information and navigation aids. Studies on pedestrian navigation pilot systems and tests with commercially available pedestrian navigation services on Smartphones have shown that key requirements like mobile multimodal trip planning or navigation in interchange buildings are not addressed adequately.

Our approach identified basic requirements for a personal travel companion and addressed key research questions concerning mobile trip management, detailed pedestrian route calculation for interchange situations in public transport interchange buildings, navigation and guidance along pedestrian routes and continuous positioning. In the implementation section we described some implementation details.

For the evaluation of the prototype we conducted a user test setting in the Vienna underground tram station Matzleinsdorfer Platz in autumn

2005. Travellers were guided by the prototype application along 6 pre-defined test routes. 20 test persons participated in the test setting and completed the user survey. The average assessment of the prototype application was 2.43 on a scale from 1 (best) to 6 (worst). 4 persons found the travel companion very good, 9 persons good, 2 persons satisfying and 5 persons not useful. Criticised issues were the rather difficult input of addresses for route planning and orientation problems with the provided maps. One important result was that orientation of people can be improved by turning the digital floor plan on the mobile phone in an egocentric perspective [24]. Besides, we realised shortcomings regarding the accuracy and reliability of the Bluetooth positioning system. After extensive tests we are still convinced that Bluetooth-based indoor positioning meets our requirements. However, in the future we have to continue with fine tuning of the Bluetooth Beacon adjustments and additional research improving positioning accuracy, reliability and usability.

Acknowledgements

The work described in this paper is mainly based on results of the Austrian research projects Vienna-SPIRIT and Open-SPIRIT. The projects were partly funded by the Austrian Federal Ministry for Transport, Innovation and Technology in the research agenda I2 – Intelligent Infrastructure.

References

[1] Baus, J., Krüger, A., and Wahlster, W., 2002: A resource-adaptive mobile navigation system. In: Proc. Int. Conf. on Intelligent User Interfaces, San Francisco, pp 15-22

[2] Blip Systems, BlipNet, http://www.blipsystems.com, Denmark, Last Access 01.2006.

[3] BlueLon: BodyTag BT-002, Bluelon Headquarters, Denmark, http://www.bluelon.com/index.php?id=263, Last Access 02.2006

[4] Brunner-Friedrich, B., Kopetzky R., Lechthaler M., Pammer, A., 2001: Visualisierungskonzepte für die Entwicklung kartenbasierter Routing-Applikationen im UMTS-Bereich. In: Angewandte Geographische Informationsverarbeitung XIII, Beiträge zum AGIT-Symposium, Salzburg

[5] Casakin, H., Barkowsky, T., Klippel, A. and Freksa, C., 2000: Schematic Maps as Wayfinding Aids. In: Freksa, C. (Ed.): Spatial Cognition II, Springer Verlag, Heidelberg.

[6] Corona, B., Winter, S., 2001: Guidance of car drivers and pedestrians, Technical Report, Institute for Geoinformation, TU Vienna, Austria

[7] Darken, R.P., Allard, T. and Achille, L.B., 1999: Spatial orientation and wayfinding in large-scale virtual spaces II. Presence, vol. 8 (6), The MIT Press, pp. iii-vi.
[8] Dedes, G., Dempster, A.G., 2005: Indoor GPS Positioning: Challenges and opportunities. In: IEEE 62nd Semiannual Vehicular Technology Conference, September 25-28, Dallas
[9] Djuknic, G.M., Richton, R.E., 2001: Geolocation and Assisted GPS. In: IEEE Computer, vol. 34(2), IEEE Computer Society Press
[10] EU-Spirit: European Travel Information System, http://www.eu-spirit.com/, Last Access 01.2006
[11] Feldmann, S., Kyamakya, K., Zapater, A. and Lue, Z., 2003: An indoor Bluetooth-based positioning system: concept, implementation and experimental evaluation. In: Proc. ICWN'03, Las Vegas
[12] Fontaine, S. and Denis, M., 1999: The production of route instructions in underground and urban environments. In: Freksa, C. and Mark, D. (Eds.): Spatial Information Theory, International Conference COSIT'99 Proceedings, Lecture Notes in Computer Science 1661, Springer Verlag.
[13] Gartner, G., Frank, A. and Retscher, G., 2004: Pedestrian Navigation System for Mixed Indoor/Outdoor Environments - The NAVIO Project. In: Schrenk, M. (Ed.): CORP 2004 and Geomultimedia04. Proceedings of the CORP 2004 and Geomultimedia04 Symposium, 2004, Vienna, pp. 165-171.
[14] Gibson, J.J, 1999: The Ecological Approach to Visual Perception. Lawrence Erlbaum Assoc., London.
[15] Golledge, R.G., 1999: Human wayfinding and cognitive maps. In: R.G. Golledge (Ed.): Wayfinding behavior, John Hopkins University Press, Baltimore, pp. 5-45.
[16] Heye, C., Rüetschi, U.J. and Timpf, S., 2003: Komplexität von Routen in öffentlichen Verkehrs-systemen. In: Proc. Angewandte Geographische Informationsverarbeitung XV, Salzburg.
[17] Hightower, J. and Borriello,G., 2001: Location systems for ubiquitous computing. IEEE Computer, vol. 34(8), pp. 57–66
[18] ISCOM, 2002: Data models and data exchange formats, IST Project Deliverable D2, Munich
[19] Johnson, M, 1987: The Body in the Mind, Univ. of Chicago Press, Chicago
[20] Krüger, A. et al., 2004: The connected user interface: realising a personal situated navigation service. In: Proc. Int. Conf. on Intelligent User Interfaces, Funchal, Portugal, 2004, pp. 161-168
[21] Leonhardt, U., 1998: Supporting Location-Awareness in Open Distributed Systems. PhD thesis, Department of Computing, Imperial College, London
[22] Lesswire: LocalNavigator, Lesswire AG, Germany, http://www.lesswire.de/, Last Access 02.2006
[23] May, A.-J., Ross, T., Bayer, S.H. and Tarkiainen, M.J., 2003: Pedestrian navigation aids: information requirements and design implications. Personal Ubiquitous Computing, vol. 7, pp. 331f

[24]Meng, L., 2005: Ego centres of mobile users and egocentric map design. In: Map-based Mobile Services, Theories, Methods and Implementations, Springer Berlin, pp 87-105
[25]Mentz, H.J.: IJP (intermodal journey planner), Mentz Datenverarbeitung GmbH, http://www.mentzdv.de/www/en/produkte/efa.htm, Last Access 02.2006
[26]Montello, D.R.: Navigation. In: Miyake, A. & Shah, P. (Ed.), 2005: Cambridge handbook of visuospatial thinking. Cambridge Univ. Press, Cambridge.
[27]Peterson, B., Baldwin R. and Kharoufeh, J.P., 2004: A specification-compatible Bluetooth inquiry simplification. In: Proceedings of the 37th Hawaii International Conf. on System Sciences (HICSS), Waikaloa, Hawaii
[28]Retscher, G. and Thienelt, M., 2004: NAVIO - A Navigation and Guidance Service for Pedestrians. In: Papers presented at the GNSS Sydney 2004 Conference, December 6-8, 2004, Sydney, Australia, 12 pgs.
[29]Rüetschi, U.J. and Timpf, S., 2004: Schematic geometry of public transport spaces for wayfinding. In: IfGI prints, Münsteraner GI-Tage, vol. 22, Muenster, 2004, pp. 191-203.
[30]Rüetschi, U. J. and Timpf, S., 2004: Modelling wayfinding in public transport: network space and scene space. In: Proc. Spatial Cognition 2004, Chiemsee, Germany.
[31]Transport for London: Journey Planner, http://journeyplanner.tfl.gov.uk/, Last Access 02.2006
[32]Van Diggelen, F., Abraham, Ch., 2001: Indoor GPS Technology. Presented at CTIA Wireless-Agenda, Dallas
[33]Verkehrsverbund Ost-Region: Fahrplanauskunft, http://www.vor.at/, Last Access 02.2006
[34]Verkehrsverbund Ost-Region, 2002: User survey about intermodal travel information systems, Vienna, December (in German)
[35]Wang, Y., Jia, X., Lee, H.K. and Li, G.Y., 2003: An indoors wireless positioning system based on wireless local area network infrastructure. In: Proc. 6th Int. Symp. on Satellite Navigation Technology Including Mobile Positioning & Location Services, Melbourne
[36]Wayfinder: Wayfinder Navigator, Wayfinder Systems AB, Sweden, http://www.wayfinder.com, Last Access 02.2006

EASYGO – A public transport query and guiding LBS

Lilian S.C. Pun-Cheng, Esmond C.M. Mok, Geoffrey Y.K. Shea, W.Y. Yan

Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University
{lspun, lsemok, lsykshea}@polyu.edu.hk, yanwaiyeung@ieee.org

Abstract

Hong Kong is a highly urbanized city with a very complicated public transport system. Local residents and tourists rely heavily on public transport for everyday travelling. To assist people in making better use of the public transport system, a multi-modal public transport query and guiding LBS system for mobile users is developed. This system can automatically determine the mobile device position with cellular network positioning technology, and with the user-defined destination, optimal traveling routes in terms of shortest travelling time and lowest fare are suggested. An overview of this system with emphasis on the system user interface, integration of GIS and mobile positioning technology are discussed in this paper.

1 Introduction

Hong Kong is a complicated, highly urbanized city with population of about 7 million. Compared to other developed cities, the rate of private car ownership are low due to the limited land, dense population and the Government policy in regulating the volume of private vehicles on roads.

For every business day, more than 4 million people rely on public transportation system to get to their destinations. There are multiple modes available and these are mostly frequent, convenient and affordable. At present, more than 10 different public transportation companies are in operation: Mass Transit Railway (MTR), Kowloon-Canton Railway (KCR), Light Rail Transit (LRT), Kowloon Motor Bus (KMB), New World First Bus (NWFB), Minibus, tram, peak tram, feeder bus, New World Ferry, Star Ferry, Lantau Bus, taxi (red, green and blue) and so on. Information concerning some of these services might be found in the web sites of these companies. Most are on assisting travellers to select a suitable route to their destination. Yet, all are concerned with a single-mode, whether be a direct route or one involving transfer. There is lack of information about the connection of other forms of transport to the destination. The choice of origins and destinations are usually defined as a larger geographical area, such as that of a district. There is also no information on how to actually get to the stops or terminals from the boarding place and/or at the connection points. To many transport users, a good and complete public transport information system should possess the following characteristics. They should be provided with information about different forms of public transport and their transfer details to the destination. They should also be able to advise the routes, transfer details and alternatives, if any, according to their preference such as "the most economic route", "the most comfortable route", "the fastest route". In addition, such information should be presented in a user-friendly environment, in the language(s) commonly adopted in that society, and in the form of easily perceived maps, charts and/or simplified texts. With this insight, the Department of Land Surveying and Geo-Informatics of the Hong Kong Polytechnic University has developed a multi-modal public transport query and guiding LBS system namely EasyGo. The project started in early 2002 and its objectives are:

- To design algorithms for a multi-modal route searching model;
- To create a web service such that results of the route searching model might be displayed and accessed by the general public for planning their trips; and
- To enable query and dissemination of such information primarily through wired and wireless communication channels.

The wireless version suitable for WAP, GPRS and 3G device is introduced below.

2 An overview of EASYGO LBS

Figure 1 shows an overall architecture of EasyGo. The system is designed in a way that flexibility is given for wired and wireless applications. In designing the wireless version of this system, the entire route searching process can be accelerated if the mobile user's position is the starting point of the journey. This is achieved by determining the mobile device location by mobile positioning technology. A location request is first sent from the web server to the service gateway. The gateway server would then send a location task request to cellular base stations. Positional information is then sent back to the web server via internet for database and route searching processes.

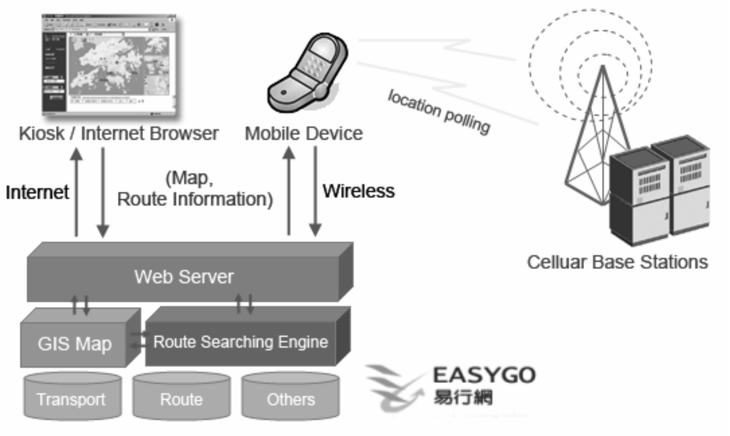


Fig. 1. Architecture of EasyGo

Figure 2 to 4 demonstrates how users can navigate the EasyGo internet system to query the public transport information.

Fig. 2. Defining Origin and Destination

Fig. 3. Route Searching Results

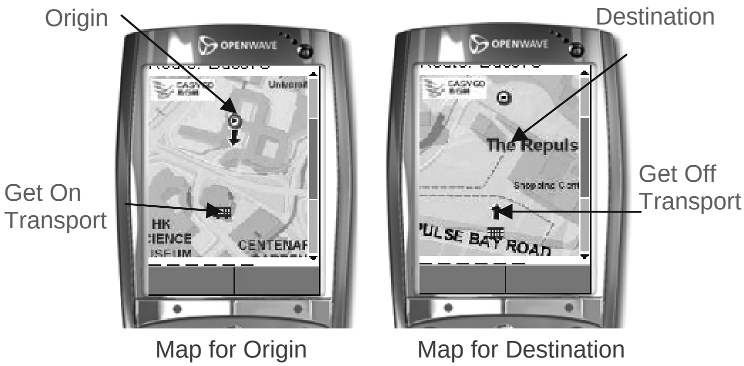


Fig. 4. Map Display Options

To start with, a user has to specify an origin and a destination in either Chinese or English. The origin (O) can be enabled by the following three ways: (i) automatically provided based on result of cellular network positioning, (ii) text search, (iii) clicking to a position on the map. Similarly, the destination (D) can be enabled using methods (ii) and (iii). According to any pair of O-D, the system will proceed to compute and recommend a summary of possible routes as shown in Figure 3. If the user is still not certain of the get on, get off and transfer locations, he may view the map by just clicking to the icon beside the pick up, get off and transfer stops in Figure 4. It should be noted that 'stop' is a general term applying to any bus/minibus/tram stops, ferry piers and mass transit station exits.

3 EASYGO Subsystems and route searching

The multi-modal travel information searching engine consists of three bilingual subsystems for enhancing the web user interfaces. These will be discussed sequentially as follows.

3.1 The Public Mode Subsystem

This contains two main databases:

1. database storing all the routes with direction of all modes and their supplementary information such as fare, frequency and conditions;
2. geographic database storing the positions of all stops and exits up to the accuracy of the pavement section these reside on.

This information come from the various bus and mass transit companies, the Transport Department, city guide map books as well as on-site checking and verification.

3.2 The Address Subsystem

It is a database of names for user-selection of origins and destinations. This includes all street names, site names (e.g. a certain residential district, a hospital) and important building names and landmarks, most of which are based on the G1000 series of the Land Information Centre, Survey and Mapping Office of the Lands Department. Similarly, a geographic database containing their positions is also developed.

3.3 The Map and Chart Subsystem

Edited from the 1:5000 base maps of Survey and Mapping Office, two series of seamless maps - 1:3000 and 1:7500 covering the whole area are prepared for user-selection of origins and destinations on the map and for displaying the route search result. These maps are designed in the way that only significant address and public mode information are highlighted. Another form of displaying the result of taking a direct or transfer routes is the provision of a chart. It lists out all the stops for the suggested routes and each stop is linked to the map showing the stop position as well.

Based on the three subsystems, the Public Transport Query and Guiding System is developed with three major functions or environments.

3.4 Computation of Optimal Routes

Once an origin and destination have been decided, the system will go on to search a matching public transport mode and route with least and preferably no transfer as possible. Proximity (that is of acceptable walking distance) to these stops from the origin and to the destination is a major searching criterion. In many cases, more than one means is available. The suggested list will prioritize the choices with the following preferences:

- direct routes with no transfer
- a mass transit route plus one transfer
- least cost of price or time taking into consideration the walking time to stops, the waiting time, the average speed of the mode, the number of stops for the traveled route and the estimated distance of the travel.

Optimal routes can be found by applying certain network analysis algorithms to the transport route network such as Dijkstra's algorithm (Dijkatra, 1959), k-quickest simple path (Chen, 1994), 2-stage heuristic method (Li & Kurt, 2000) and time and cost minimization (Ahuja, Orlin, Pallottino & Scutella, 2003). A detailed analysis of different shortest-path algorithms may also be found in Pallotinno & Scutella (1998) and Zhan & Noon (1998). In this application, the public transport route network for the whole Hong Kong territory is so large in terms of node and edge numbers that computation becomes inefficient. Therefore, some measures must be taken to limit the range of searching for possible travel routes. Several techniques have been used to improve the route-searching performance and shorten the response time. One is to use two-stage route planning/searching as a compromise of algorithm complexity and computation efficiency. Briefly speaking, the global and local routes are planned at different stages. The whole global network might be partitioned into smaller local networks within Hong Kong Island, Kowloon, New Territories and Lantau Island, nesting on these are the administrative divisions of 18 districts. When a request is submitted, the system will first examine if the origin and destination are within the same partition and search routes in the respective local network if they do. If otherwise, searching has to be referred to the global network but it must be reduced. In doing so, the mass transit routes such as MTR, KCR will be considered first, followed by bus routes and ferries. This can reduce a large number of minor or trivial routes within the various partitions in concern.

The derived information (i.e. the suggested routes) will have to be presented to users clearly and easy-to-understand. This is in the form of a chart complemented with textual descriptions (Figures 2 to 4). Not only all the stop names connecting the origin and destination are shown, but also each stop is linked to a map showing its location.

4 Accuracy of mobile positioning

As mentioned in 3.4 above, proximity to transportation stops from the origin and to the destination is a major searching criterion. If mobile positioning is used for determining the traveling origin, accuracy is a key factor for providing appropriate route advice to LBS users. It is believed that cellular network positioning based LBS platforms now available in Hong Kong can generally provide positional information better than 300m accuracy in urban area, which would be sufficient for transportation query and route guiding. An investigation was therefore carried out to verify the accuracy achievement of current LBS in Hong Kong.

4.1 System Development for the Accuracy Test

A location polling system containing location information retrieval and map display functions was developed using ESRI's GIS MapObjects. As can be seen in Figure 5, location information was obtained through bi-directional communications between the location polling system, service provider's location gateway, and the mobile device installed with a special JAVA SIM card. The location polling system sends location task request to base station by passing the service provider gateway through Internet. Location information is returned back to the service provider and passed back to the location polling system with GIS display in XML file format. The XML file includes mobile ID number, northing and easting coordinates, street name, date and two accuracy indexes namely Cell-ID Accuracy Index and Network Measurement Result Accuracy Index (NAI).

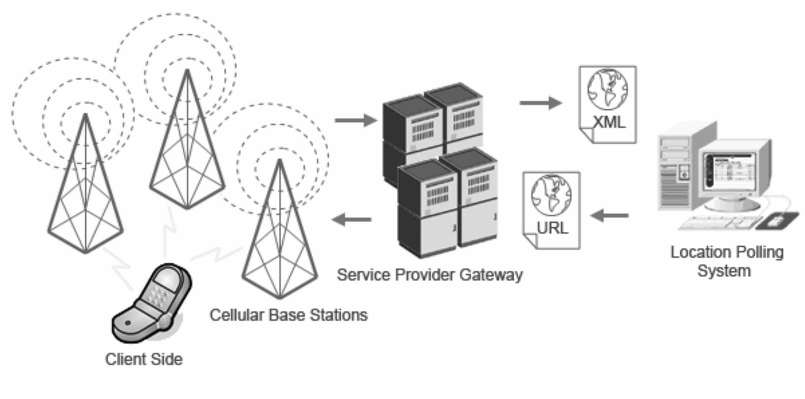


Fig. 5. System Design of the LBS Test

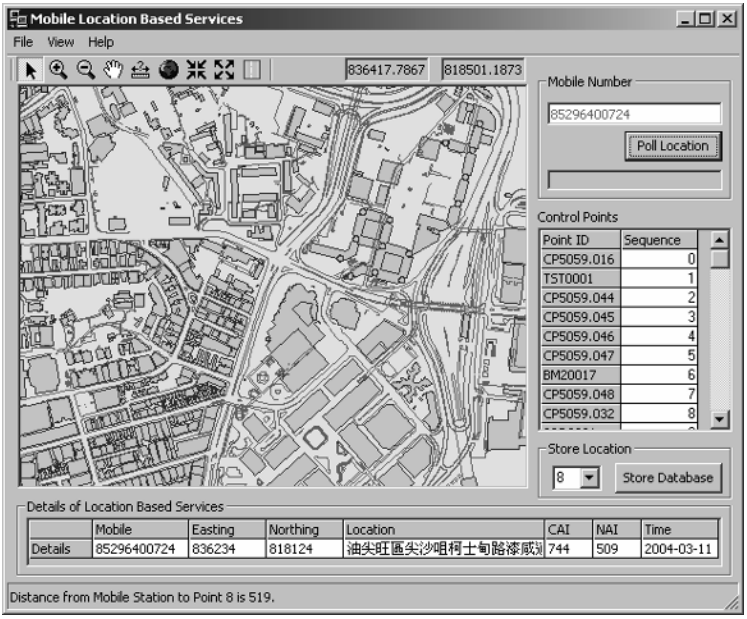


Fig. 6. Location Polling System interfaced with the GIS

The location polling system described above was used to verify the accuracy achievement and stability of cellular network positioning along Nathan Road and Sai Sha Road, representing both urban and rural areas. Nathan Road is one of the busiest areas in Hong Kong surrounded by dense high rise buildings, whereas Sai Sha Road is a narrow two-way road connecting the less dense Ma On Shan and Sai Kung areas. In this test, pre-selected positions with known coordinates were used to compute the positioning accuracy. These positions include existing government survey marks and well defined features which can be clearly identified on digital maps. A total of 82 points along Nathan Road, and 49 points along Sai Sha Road were selected.

5 Accuracy Assessment Results

Positioning error results are analyzed with ArcView GIS, and the field test results are summarised in Table 1. Figures 7a to 7b show the distribution of pre-selected points and the generated error buffer along Nathan Road and Sai Sha Road. For the test along Nathan Road, the minimum error is 26m and the maximum error is 400m. For the test along Sai Sha Road, the minimum error is 76m while the maximum error is 2662m. The statistics on different error range are shown in Figures 8a and 8b respectively.

Table 1: Summary of Field Test Results

	Nathan Road	Sai Sha Road
Length	Around 3.5 km	Around 5.7 km
Pre-selected Point	82 points	49 points
Maximum Error	400 m	2662 m
Minimum Error	26 m	76 m
Root Mean Square Error	180 m	1202 m

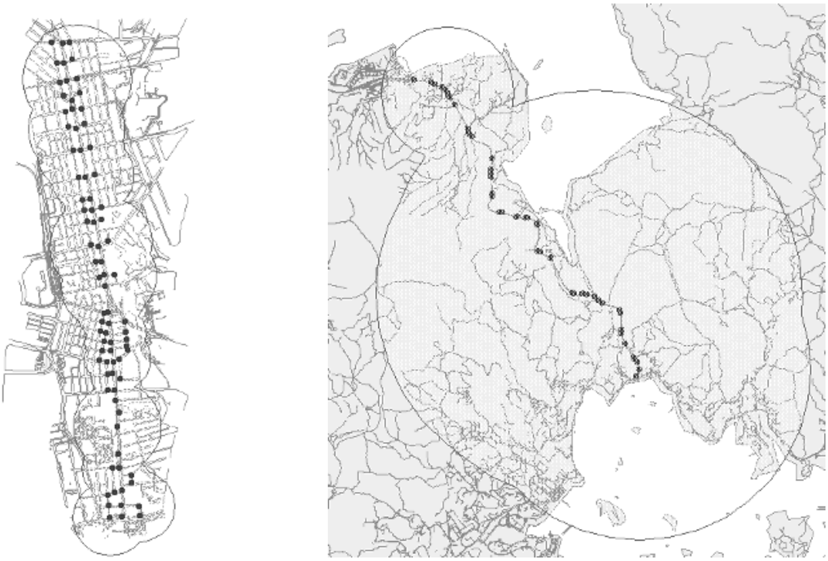


Fig. 7. a. Error Buffer along Nathan Road, b. Error Buffer along Sai Sha Road

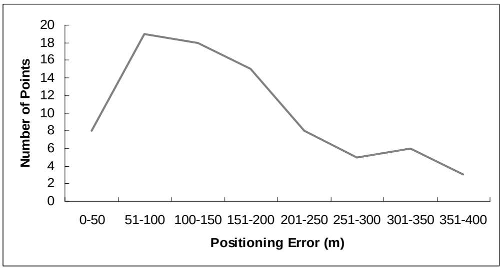


Fig. 8a: Error Statistics (Nathan Road)

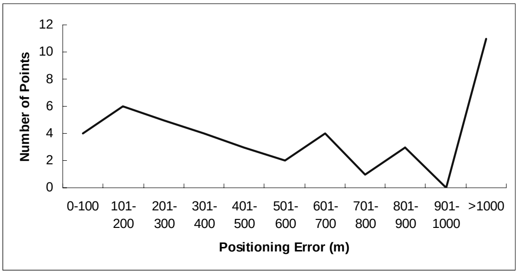


Fig. 8b: Error Statistics (Sai Sha Road)

6 Concluding remarks

With a view that a multi-modal transportation system has so far not been investigated in the mobile market but is deemed necessary for the general public, a system has been developed with strong inputs from digital mapping and geographic information system. In summary, this system is characterized by:

a. automatic user position determination option to facilitate setting of traveling origin;
b. a truly *point-to-point* route searching engine at street block detail level;
c. one-stop route searching that accommodates all transport modes and every transport route;
d. intelligent route searching which takes into account walking distance and multi-modal transfer; and
e. being user-friendly especially when navigating the multi-scale maps and various flexible user interaction methods

With regard to geolocation service in Hong Kong, according to the field tests, cellular network positioning can generally perform better than 200m from actual location in urban area and the error can be bigger than 1km in rural area. Accuracy of the mobile position fixing depends on many factors, such as the number of cellular stations and cell size in the network, geometric configuration of cellular network, signal strength, and positioning method. For example, the Time of Arrival (TOA) and Time Difference of Arrival (TDOA) positioning methods determine positions with distances between cellular stations and the mobile device, which require the signal transmission and reception time information. However, blockage; diffraction and reflection of signals propagating through different environments would cause the Non-Line-of-Sight (NLOS) time delay error. This error would lead to the error in distance determination and would eventually cause serious positioning error. To increase the accuracy of LBS in Hong Kong, one may argue that GPS alone system can generally achieve better than 20m accuracy, which is far better than the current cellular positioning methods. It should be noted that, the GPS alone method is very susceptible to signal obstructions and multipathing, whereas in dense high rise environments, successful and accurate position determination will become much more difficult, if not impossible. The combination of GPS and cellular network positioning can solve the problems of blind network coverage and unreasonable results in rural areas, since most New Territories areas are quite open to the sky.

The trend for LBS in Hong Kong, from the point of view of the authors, will gradually move to the hybrid positioning technique, which is the combination of AGPS and cellular network positioning (Mok and Xia, 2006). One of such systems namely gpsOne™ developed by Qualcomm has been applied in CDMA networks, to provide high accuracy indoor and outdoor location based services (Qualcomm, 2004). Moreover, low-cost, high speed communication and positioning enabled systems such as Wi-Fi (Patil, et al., 2005) and Ultrawide Band (Schroeder, Galler and Kyamakya, 2005) technologies are emerging the Hong Kong community and they are seen to have great potential for developing 3-dimensional and high accuracy ubiquitous location based services, to support a varieties of location based services applications.

Acknowledgement

The maps presented in the location polling system are generated based on the digital map data provided by the Land Information Centre of the HKSAR and the work presented in this paper is supported by the Research Grant of the Hong Kong Polytechnic University Project No. G-T29B and G-T452. Special thanks are also given to Miss Frances Wan Ka-wing, Miss Tang Yuen-fun and Miss Lee Sze-man for their help in field test.

References

Ahuja, R., Orlin, J.B., Pallottino, S. & Scutella, M.G. (2003) Dynamic shortest path minimizing travel times and costs, Networks, vol. 41(4), pp.197-205.

Bosak, J. (1997) XML, Java, and the future of the Web, Sun Microsystems Technical Paper.

Centamap, Hong Kong http://www.centamap.com

Chen, Y.L. (1994) Finding the k quickest simple paths in a network, Information Processing Letters, vol.50, pp.89-92, Elsevier.

Dijkstra, E.W. (1959) A note on two problems in connection with graphs, Numeriche Mathematik, vol.1, pp.269-271.

ESRI (1998) ArcView internet map server extension: functional overview, an ESRI White Paper, Environmental Systems Research Institute Inc.

Gifford, F. (1999) Internet GIS architectures – which side is right for you? In GIS World, vol. 12, no. 5.

Gunn, M. (1999) Squarely at the corner of MapInfo's vision: Taking leadership in Internet-based mapping solutions, in MapWorld Magazine, vol. 4, 1, pp. 10.

Kowloon Motor Bus Company.(KMB), http://www.kmb.com.hk

Kunczier, H. and Kuchar, A. (2002), Network measurement report based handset localization. White Paper of Telecommunication Research Center Vienna, Vienna, Austria, 20 June 2002.

Li, Q. & Kurt, C.E. (2000) GIS-based itinerary planning system for multimodal and fixed-route transit network, Mid-Continent Transportation Symposium Proceedings, pp.47-50.

Mok, E.C.M., Shea, G., Yan, W.Y. (2004) Geolocation Positioning with Wireless Cellular Network in HongKong, The HongKong Institute of Surveyors Journal, Vol.15, 2, pp.23-30.

Mok E.C.M. and Xia L.Y. (2006) Hybrid GPS and Wireless System for Geolocation Positioning in Urban Canyons, Proceedings of International Workshop on Successful Strategies in Supply Chain Management, 5-6 January 2006.

New World First Bus Company, Hong Kong, http://www.nwfb.com.hk

Pallotinno, S. & Scutella, M.G. (1998) Shortest path algorithms in transportation models: classical and innovative aspects, in Marcotte, P. & Nguyen, S. (eds.) Equilibrium and Advanced Transportation Modelling, Kluwer Academic Publishers, pp. 245-281.

Patil, A., Munson, J., Wood, D. and Cole, A. (2005) BlueBot: asset tracking via robotic location crawling, Proceedings of International Conference on Pervasive Services, 11-14 July 2005, pp.117 - 126.

Pun-Cheng, L.S.C. & Shea, G.Y.K (2003) Design parameters for an interactive web mapping system, in Li, Z., Zhou, Q. & Kainz, W. (Eds) Advances in Spatial Analysis and Decision Making, pp. 299-306, A.A. Balkema Publishers.

Qualcomm CDMA Technologies (2004), gpsOne™ Overview, http://www.cdmatech.com/solutions/pdf/gpsone_overview.pdf

Schroeder, J., Galler, S. and Kyamakya, K. (2005) A Low-Cost Experimental Ultra-Wideband Positioning System, Proceedings of IEEE International Conference on Ultra-Wideband, 05-08 September 2005, pp.632 - 637.

Zhan, F.B. & Noon, C.E. (1998) Shortest path algorithms: an evaluation using real road networks, Transportation Science, vol. 32, no.1, pp.65-73.

Mobile City Explorer: An innovative GPS and Camera Phone Based Travel Assistant for City Tourists

Siegfried Wiesenhofer[1], Helmut Feiertag[1], Markus Ray[1], Lucas Paletta[2], Patrick Luley[2], Alexander Almer[2], Mathias Schardt[3], Josef Ringert[4], Paul Beyer[5]

[1] arsenal research, Vienna, Austria, siegfried.wiesenhofer@gmx.net
[2] Joanneum Research, Institute of Digital Image Processing, Graz, Austria
[3] Graz University of Technology, Institute of Remote Sensing and Photogrammetry, Graz, Austria
[4] Graz University of Technology, Institute of Geoinformation,Graz, Austria
[5] three united mobile solutions, Vienna, Austria

Abstract

We present a concept and architecture to provide location aware travel assistance for city tourists based on a combination of wide spread conventional mobile phones with built in cameras and GPS/GALILEO receivers. Novel approaches to improve the support of tourists pre-, on- and post-trip are outlined. The Mobile City Explorer (MCE) enables city tourists to start a tour without time-consuming planning tasks. It identifies sights by taking photos, leads tourists to sights matching their interests and lets them collect multimedia impressions with their mobile phone along the tour. The travel diary stores the collected impressions, is accessible over the Internet, and allows editing and sharing with family and friends. In this paper we outline the system architecture and components that enable these new approaches.

1 Introduction

A city tourist experience often begins with much preparation work before the tourist actually arrives at the desired destination (pre-trip phase): The tourist gathers information about the chosen destination and plans which sights to see, either before departing from home or during travelling to the destination. He might read a travel guide or search for detailed information on the Internet to identify points of interest (POIs) he wants to visit, prioritize the POIs and create an itinerary to follow.

On-trip, the tourist usually starts following his planned itinerary to a certain extent. Occasionally, he spontaneously deviates from the planned route and explores additional routes and POIs when seeing other interesting objects (on-trip phase). He tries to find routes and POIs on a map, and physically locates and identifies sights. The tourist frequently stops at certain sights to take pictures or to read related information in a guide he is carrying during the trip. Some travelers like to jot down their impressions in a travel diary, either on trip or at night in a restaurant or their hotel room.

Finally, after having arrived back home (post-trip phase), some travelers sort out the large amount of pictures taken and eventually put them into a commented diary in order to share the experiences with family and friends. Many people, however, do not find the time to get their pictures and travel notes organized and are not able to recall the sights on the pictures a few years later.

Mobile City Explorer (MCE) is a project implementing the concept of an innovative mobile guide for personalized city tours. The concept was derived from the user perspective outlined in the above scenario, integrating features for guiding the tourist according to his preferences and actual behavior, identifying objects using object recognition, and collecting pictures and route information in an automated travel diary. The project objective is to develop a location aware pilot system for city tourists employing camera phones and GPS/GALILEO positioning technology. The mobile travel assistant guides tourists through the city, suggests POIs matching the tourists' interests and recommends appropriate routes. If the tourist deviates from the recommended route, the tour is recalculated according to his personal interests and time constraints. Using object recognition the pilot system is able to identify POIs from pictures taken with the camera phone and provides the relevant information to the tourist. Figure 1 illustrates this application scenario. At the same time the personal multimedia travel diary collects pictures, videos, text comments and acoustic impressions captured by the user along the route. This saves the user from the tedious work of documenting, cataloguing, and processing his pictures and enables him to “relive“ his travel experience post-trip with family and friends.

Fig. 1. Mobile City Explorer Application Scenario

In order to support a wide adoption in the market, a thin client architecture using digital camera phones and GPS/GALILEO receivers has been chosen for the mobile client. ABI Research estimated that almost 70% of all mobile handsets will have an integrated camera by 2009. Also, market forecasts predict 290 million phones with integrated GPS/GALILEO modules worldwide by 2009. It can be expected that the increasing market for location based services, GALILEO enabled improvements in positioning precision and regulations such as US911 and E112 will further propel the supply and demand of mobile devices with integrated GPS/GALILEO modules.

The rest of this paper is organized as follows. Section 2 sketches typical application scenarios of a city tourist using the MCE. The system architecture and components to realize such a system are explained in section 3. Finally, section 4 summarizes our work.

2 Application scenarios

In this section we present the essential application scenarios to illustrate the MCE features. Figure 2 provides a schematic overview of the functionality of the MCE system. To use the service a mobile phone with a WAP browser and an external GPS/GALILEO receiver is required.

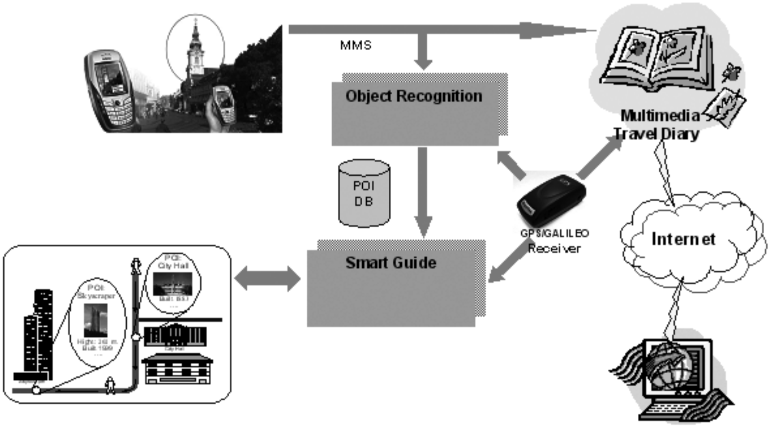


Fig. 2. Schematic view of the MCE functionality

2.1 Service registration

Initially, a tourist registers for the service via a WAP or Web browser. During the registration process, he chooses a login name, password and provides his phone number and other personal data. After having completed the registration, the tourist can download and install an application via WAP by a provided URL. The application implements the positioning procedure using the GPS/GALILEO positioning system and forwards the positioning data to the MCE server for further analysis.

2.2 Starting a personalized tour

After having started a new tour, the tourist logs into the MCE service with her mobile phone using the WAP browser. Before he starts her first tour he fills in his interests in an online questionnaire. Subsequently, the MCE smart guide determines an initial sequence of POIs matching the specified interests. The suggested sequence of POIs takes into account the user interests, current position and time constraints. This makes it possible to provide an individual trip best matching a tourist's interests such that it is possible to visit all POIs within a predefined time span, e.g. a day.

Then a WAP Push message is sent to the tourist, where he receives a specially formatted SMS that displays an alert message about the new route and gives the option to connect directly to a URL via the mobile phone's WAP browser. The URL leads to a WAP page, shown in figure 3, including a map showing the current position of the tourist marked with a dot in the center of the map and the nearby POIs marked with a sequence number presenting the suggested route, illustrated in figure 3a. Additionally, hyper links to pan and zoom the map, see figure 3b, and a hyper linked POI list to detailed information of all POIs within the map are provided in the WAP page, as shown in figure 3c.

(a) (b) (c)

Fig. 3. (a) WAP page with map (b) links to pan and zoom (c) a linked POI list

2.3 Guiding the tourist and route adaptation

On the mobile phone, a client application regularly transmits position data to the MCE database. The transmitted position data are analyzed by the MCE smart guide. When the tourist comes close to a POI of the suggested route, the MCE smart guide sends a notification via a WAP Push message to the tourist's mobile phone. An example for a POI notification is shown in figure 4. The mobile phone displays an alert message, as shown in figure 4 (a) with the name of the POI "Staatsoper" and gives the option to connect directly to a particular URL to get detailed information about that POI. After opening the URL the WAP browser will show the WAP page with detailed POI information, figure 4 (b), and the user can scroll down to a detailed description, figure 4 (c), and further images, figure 4 (d).

a) b) c) (d)

Fig. 4. (a) POI notification via WAP Push message displaying the alert message (b) showing the WAP page (c) and scrolling down to detailed description (d) and image

The MCE collects feedback about the motion behavior of the tourist according to how much time he has spent at the particular POI and whether he has followed the link to the detailed POI information. The service learns from the motion and interaction history of the tourist and suggests a route according to the relevance of the POIs not yet visited. If the tourist deviates from the suggested route, the route is adapted by inserting nearby POIs matching his interests and removing POIs beyond defined distance and time limits. A WAP Push message is sent to the tourist containing a link to a WAP page, which includes the recalculated sequence of POIs and the corresponding map.

2.4 Collecting impressions

Tourists can collect impressions on-trip by writing text, recording audio comments and taking photos and videos via the built in camera and send them via Multimedia Messaging Service (MMS) to the MCE service. The server extracts the multimedia impressions from the received message and stores them in a new entry in the travel diary including time stamp and geographical position. Each MMS results in a travel diary entry. The geographical position for that entry is retrieved from the tourist's motion history using the transmission timestamp of the MMS.

2.5 Recognizing sights by taking photos

Tourists can identify POIs by taking a picture of a sight and sending a MMS to the MCE object recognition component. On successful recognition a WAP Push message is returned to the tourist's mobile phone containing the name of the recognized sight and a link to a WAP page with detailed information. If the recognition fails, e.g. due to poor image quality or lack of details in the image, a SMS with an error message is returned.

2.6 Revising and sharing travel impressions

Back home from the travel the tourist is able to log into the MCE multimedia travel diary via a Web browser, edit his entries and send an Internet link to his travel diary in order to share his impressions with family and friends post-trip.

3 System architecture and components

In this section we present an overview of the system architecture and components we use for the MCE. The functional responsibilities of the MCE are mapped onto components as illustrated in figure 5. The following sections provide details for each of the main components.

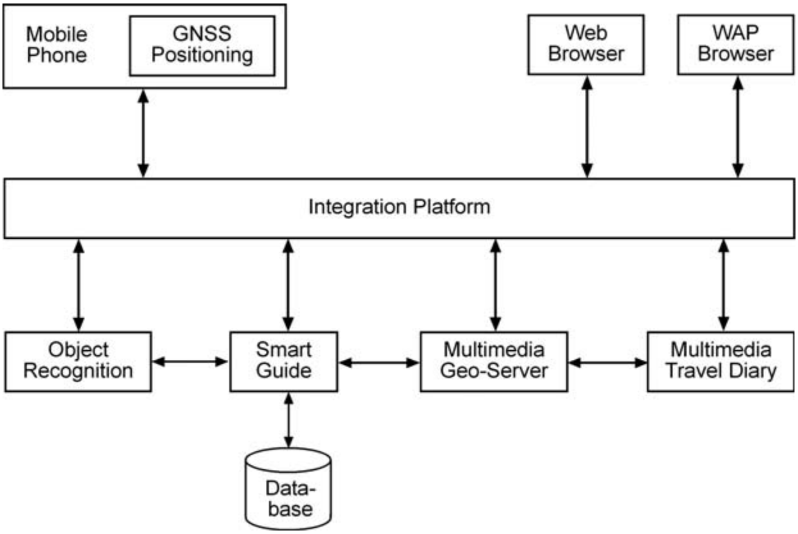


Fig. 5. MCE system components

Section 3.1 explains the smart guide which observes and learns interests from tourists' behavior and makes route suggestions based on those interests. In section 3.2 the multimedia geo-server is introduced supplying the cartographic representation for orientation and information. Section 3.3 introduces the object recognition which recognizes sights from photos taken with the camera phone. The multimedia travel diary is explained in section 3.4. Finally, section 3.5 gives a short description of the integration platform.

3.1 Smart Guide

The smart guide operates as an electronic travel companion guiding the tourist to places and sights along tours which are suggested based on pre-trip selection and on-trip learning of user interests (e.g. architecture, museums, shopping, etc.).

To provide such a High Level Routing (HLR) the following steps are necessary:

- Collect the feedback from user’s behavior.
- Statistically evaluate the user’s interests based on the feedback from user’s behavior.
- Calculate a route by compiling a set of suggested POIs matching the user’s interests.

3.1.1 Feedback collection of the smart guide

Feedback of the user’s behavior is collected from two different sources: First the movement of the user is monitored with a Java application installed on the mobile phone and retrieving and sending positioning data from a GPS/GALILEO receiver to the MCE database via GPRS. The motion based event detection detects events such as the visit of a particular POI which leads to adaptations of the user model. Secondly the interaction with the information supplying application is observed. Thus information requests about a particular POI also lead to changes in the user model.

3.1.2 Feedback evaluation of the smart guide

The central concept for the HLR system is the so called user model. The user model includes a sociodemographic part and a part mapping the user’s interests. The interests are hierarchically ordered and for each user an individual inclination to each interest is stored (represented as a number in [0,1] where zero corresponds to no interest and one to highest interest). In order to obtain a suitable initial user model, an explicit acquisition of user interests is performed using a brief initial online questionnaire during registration. This way personalization is guaranteed at the beginning of the first session with HLR.

The learning of user interests is based on relevant feedback (like e.g. which POIs users have visited) which is on the one hand extracted from the motion behavior and on the other hand related to which POIs users have requested information on. This information is evaluated statistically for each user’s interest using an algorithm for univariate significance analysis developed and evaluated by Schwab and Pohl [SP99]. The learning algorithm adapts the current user model with regard to the new information according to the feedback and stores the adjusted probability of each interest in the individual user model. In more detail the adaptation according to visited POIs uses a classification of the POIs: To each POI a profile of interests (indicating to which interests the POI has relevance) is manually assigned. If a user visits a specific POI and requests detailed information about that POI, then it can be assumed that the user is interested in the POI and therefore he will be interested in POIs matching the same profile of interests. The user model is updated accordingly.

3.1.3 Route calculation of the smart guide

Route calculation deals with the selection of POIs that matches the interests of the user and the calculation of a route including the maximum number of POIs possible to visit in a given time, starting at the current position of the user.

The challenge is to find POIs best matching a given user model. For that a common basis is necessary and thus to each POI a profile of interests (see above) is assigned. Note that a POI can correspond to several interests. The similarity between user interests and a given POI is determined by computing the Spearman correlation coefficients [FK02]. Subsequently POIs are ranked according to their corresponding Spearman correlation coefficients. The suggested POIs are then selected from the top ranked POIs also taking the walking distance to reach the POIs into account. On trip the user interests are recalculated from the user modeling system, shown in figure 6, using feedback from motion behavior and information retrievals. Changes in interests will cause a recalculation of the Spearman correlation coefficients of the POIs and furthermore result in a change in the suggested route.

3.1.4 Schematic overview of the smart guide High Level Routing

Figure 6 shows the control loop of the HLR in a schematic way. First the user model is initialized using the information on the individual interests collected during the service registration. Afterwards the service predicts interests and preferences for the individual user based on those of similar users and group membership of the user. Based on this user model and taking the current position into account, a tour of POIs is compiled and sent to the user. The user starts his tour and is supervised by the HLR service which learns his interests and preferences based on his motion behavior and information retrievals. Deviations from the suggested route will cause updates in the user model, which in turn leads to reselection of POIs and tour recalculation.

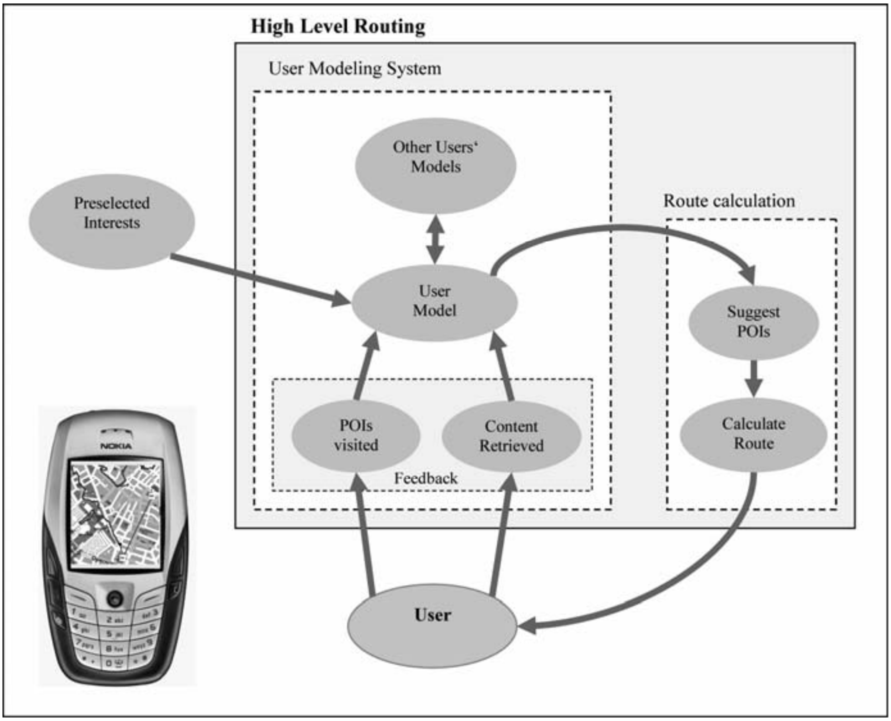


Fig. 6. Schematic view of guiding a tourist with High Level Routing

3.2 Multimedia Geo-Server

The multimedia geo-server provides multimedia tourist information supplemented by a cartographic representation of the surrounding of the tourist's current position. This is vital for the orientation and information of the tourist on-trip and also for the presentation of the visited sites post-trip. The cartographic representation is rendered for devices with small displays.

The multimedia geo-server is based on the MapServer originally developed by the University of Minnesota in the version 4.6. The server platform itself is a Linux system. For the dynamic and interactive generation of digital maps, according to the different user scenarios defined within the project, the programming language "PHP 5.0" is used. The open source PHP-extension "MapScript" enables direct access to the functionality of the MapServer, which makes the development of an effective map generating application possible. All user specific data, which need to be visualized in the map, are retrieved at runtime from the MCE databases. This enables the representation of up to date data, like the current position of the tourist, in the map. For each different type of map a specific method is developed, which can be called from all other modules externally. [LPASR05]

The component is responsible for:

- Generation of tourist maps to be displayed on the mobile device
- Generation of tourist maps to be displayed in the Internet
- Generation of a map with different route recommendations
- Conversion of an address into geographic coordinates, in case positioning using satellites is not possible

A tourist map has the following content:

- A city map from Freytag & Berndt as base map
- The reclined track of the tourist
- The current position of the tourist
- The high-level routing information, which enables the tourist to follow the track
- POIs along the track. The sights are visualized using different symbols belonging to predefined categories

In addition to different kinds of on-trip tourist maps there is a special map type needed for the visualization of the digital diary. This map type will be used post trip to show all data collected by the tourist during a trip. The map contains the following information:

- A city map from Freytag & Berndt as base map
- The reclined track of the tourist
- The visited POIs of the tourist
- The pictures taken by the tourist for the purpose of object recognition or diary entries

In the case of generating a map to be used for the Internet presentation of MCE the geo-server additionally provides an image map in order to achieve better interactivity. The image map defines all parts of the map where a POI is shown and provides a URL for each of these areas. By clicking these areas the user can follow the URLs and view detailed information about the selected POI.

In all cases of map generation the geo-server additionally provides a list of all POIs within the map containing the name of the POI and the corresponding URL for viewing detailed information about it. In addition a list containing geographical coordinates, which are defining specific map areas (bounding-boxes), used for navigation purposes on the client side, are calculated and provided by the geo-server.

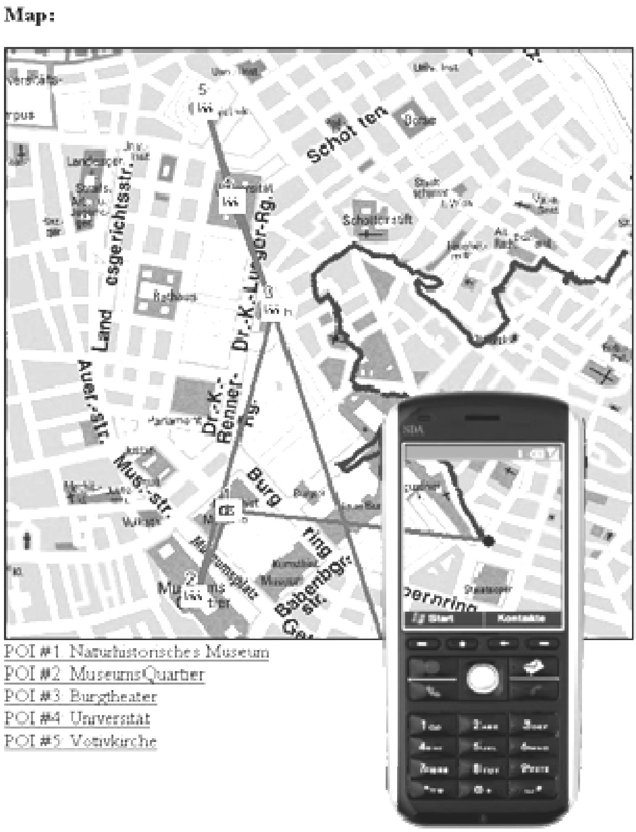


Fig. 7. Example of a map generated by the geo-server component

3.3 Object Recognition

The object recognition component recognizes POIs, i.e. tourist relevant objects like sights, objects on squares, and parts of buildings from the mobile imagery. The component extracts and analyzes the image attributes, and compares the results with stored attributes from geo-coded reference objects. The identity of the object is identified by probability based methods and enables a subsequent preparation of context relevant information to the tourist. As the location of recognized objects is well known, the vision module can also provide location awareness within the MCE system. The recognition component is developed in cooperation with research efforts outlined within the EU-IST project MOBVIS.

An image captured on the user side is first transferred to the server via MMS. On the server side, image analysis using a matching between image features and features residing in a POI based feature model database (FMDB) then provides a resulting POI hypothesis. Since cities can be populated with thousands of POIs, geo-contextual information in terms of the GPS signal is used to index only into locally relevant hypotheses within the complete FMDB. This makes the decision making both less complex and more successful.

For the computer vision based recognition of POIs a local feature based approach is used. The training phase consists of four major steps: (i) local feature extraction, (ii) selection of informative features, (iii) learning of rapid mappings from descriptors to hypotheses, and (iv) building of an appropriate sparse data structure for fast nearest neighbor indexing. Firstly, we calculate the SIFT descriptor [LOW04] on all training images: At local extrema of the DoG function (keypoint locations) a 128 value gradient orientation histogram is calculated. The DoG detector in combination with the SIFT descriptor is designed to be scale and rotation invariant and to some extent also illumination invariant. The SIFT descriptor is represented in a lower dimensional eigenspace determined from principal component analysis (PCA). Only informative features with respect to object classification are then retained (a local estimation of the posterior conditional entropy in feature space is used as measure of information). Only informative features are stored in the training database (on average only about 10% of all features). This *i*-SIFT approach leads to decreased search time and increased recognition performance [FSP05]. At the end of the training phase, a decision tree for fast entropy estimation and a k-d tree for fast nearest neighbor indexing are built.

For each test image SIFT feature extraction is performed. The resulting descriptors are mapped to the PCA based subspace. The feature entropy is estimated using the decision tree built in the training phase. Only the informative test image features (threshold on estimated entropy) are further processed, which reduces the number of nearest neighbor searches. From a matching between extracted and stored features, matching candidates vote for a resulting posterior distribution over object hypotheses. The recognition result is determined using the maximum confidence hypothesis, if the number of votes is above threshold and the entropy would reject uniform like distributions.

Experiments were performed on the TSV-20 database (Tourist Sights Vienna). It consists of images of about 20 tourist sights in the center of Vienna. For each object we have 2 training images (acquired with a digital camera) and 2 test images, acquired with a mobile phone (see figure 8 for sample imagery). The images are resized to a resolution of 640x480 pixel. On the test image we were capable to achieve a recognition rate of 92.5% (figure 9) based on a preliminary state of implementation.

Fig. 8. Examples of training (first and second column) and test images (third and fourth column

Fig. 9. Recognition results for objects no. 5 and 16, respectively. First and second column: training images, third column: test image with labeled SIFT features, fourth column: final distribution over object hypotheses.

3.4 Travel Diary

The multimedia travel diary is the interactive turntable of the travel impressions for the tourist. Pictures and audio clips taken with the camera phone are collected and are automatically time-stamped and geo-coded by recording the user's GPS/GALILEO location. The travel diary is accessible via the Internet and via WAP and offers the functionality of cataloguing, editing, and sharing of the stored travel impressions. This saves the user from the tedious work of documenting his travel experiences.

As usability is a vital precondition for wide spread usage, the development is focused on adequate design and structure. Usability knowledge regarding user habits, as well as lessons from best practice examples are merged, so as to guarantee the diary's utility. Further emphasis is put on which data is assigned to a user's entry. Consequently, the well integrated presentation of the map, being an essential feature of the travel diary, is of great importance. The map shows the user's location, the surrounding POIs, his path in form of a dotted line marking the points where the user's geo data had been saved. The whole map is presented as an image map in order to enable the user to directly navigate to detailed information, e. g. detailed information about a POI visited. The map's main purpose is to relate impressions and position.

Individual diary entries are listed in chronological order with the latest one on top. In the first rows of the frame, which contains the diary entry, the title, the diary name and the date are listed. Below the related text and multimedia impressions recorded and assigned to the specific log are presented. Furthermore, the user has the possibility to edit his entries post trip. While related images are presented directly in a resized format, audio and video recordings are presented via link and meaningful icons. Each features a play button, which - when clicked – opens a new window in which the file will be started. Finally a ‘Tell a friend’ button allows the user to share his impressions with family and friends.

The web based travel diary contains:

- Diary name
- Title
- Short description
- Pictures taken by the user
- Related pictures from the TIS (tourist information system) database
- Audio clips
- Video clips
- User's track
- Visited POIs
- Related time and geo data
- Map including sensitive areas linking to the related detail info

3.5 Integration Platform

The role of the integration platform is to integrate and orchestrate all the system components. For communication purposes between components within the MCE system architecture XML-RPC interfaces are implemented. The XML messages between the modules are exchanged over network connections using the hyper text transfer protocol.

The integration platform is the external interface to the end user's mobile device via the mobile network, provides services for incoming MMS messages and WAP requests as well as outgoing SMS and WAP Push messages in order to notify the user of specific events. The integration platform also handles the incoming tracking data from the users' positions and performs user management tasks such as registration and login.

4 Summary

We have presented a concept and architecture for a location aware travel assistant that learns interests from the tourists' behavior, recognizes sights from photos and collects multimedia impressions along a tour taken with the mobile phone. The Mobile City Explorer (MCE) concept supports tourists by reducing travel preparation times compared to the traditional manual way. During a trip, tourists are continuously kept informed about their position and surrounding POIs, are provided with detailed descriptions of POIs and are supported to identify POIs they face using object recognition. Finally, MCE saves the user from the tedious work of documenting, cataloguing, and processing his collected impressions and enables his to relive his travel experience post-trip with family and friends.

The novelties of the Mobile City Explorer include mobile object recognition using digital images, automated collection of impressions in a travel diary and learning personal interests from behavior. The thin client architecture in combination with mobile phones and the WAP Push mechanism was chosen in order to support a wide adoption in the market using digital camera phones and GPS/GALILEO receivers. Market forecasts predict 70% of all mobile handsets to carry an integrated camera and 290 million phones to carry integrated GPS/GALILEO modules worldwide by 2009. It can be expected that the increasing market for location based services, GALILEO enabled improvements in positioning precision and regulations such as US911 and E112 will further propel the supply and demand of mobile devices with integrated GPS/GALILEO modules.

The architecture to implement the concept for real world environments consists of the following main components: An *integration platform* provides the communication gateway, a *client application on the mobile phone* provides and collects position data, and a *smart guide* monitors tourists, learns interests from their behavior and makes personalized route suggestions. A *multimedia geo-server* supplies the cartographic representation, *object recognition* recognizes sights from captured images and a *multimedia travel diary* enables presenting and sharing the collected impressions on the Internet.

The architecture is currently being implemented. First results of a complete MCE city tour experience will be available in summer 2006.

Acknowledgements

The project is co-funded by the Austrian Ministry of Transport, Innovation and Technology (BMVIT). We thank our colleagues at arsenal research, Dietmar Bauer for his suggestions regarding statistical analysis and Norbert Brändle for his support.

References

[FK02] Fink, J. and Kobsa, A., User Modeling for Personalized City Tours, Artificial Intelligence Review, 18 (1), pp. 33 – 74, 2002

[FSP05] Fritz, G., Seifert, C., and Paletta, L., Urban Object Recognition from Informative Local Features, Proc. IEEE International Conference on Robotics and Automation, ICRA 2005, Barcelona, Spain, April 18-22, pp. 132-138.

[LOW04] Lowe, D., Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.

[LPASR05] Luley, P., Paletta, L, Almer, A., Schardt, M., Ringert, J., Geo-Services and Computer Vision for Object Awareness in Mobile System Applications, Proc. of the Symposium 2005, LBS&TC 2005, Vienna, Austria, November 28-30, 2005, pp. 61-64.

[SP99] Schwab, I. and Pohl, W., Learning Information Interest from Positive Examples. In Proceedings of UM99 Workshop on Machine Learning for User Modeling. Banff, Canada, 1999

Development of Cultural Inheritance Information System using LBS Technologies for Tourists

Byoung-Jun Kang, Yosoon Choi, Hyeong-Dong Park

School of Civil, Urban and Geosystem Engineering, Seoul National University, Seoul, South Korea
hpark@snu.ac.kr

Abstract

The object of the study outlined in this chapter is to propose a prototype of cultural inheritance information system that provides tourists with more interactive and dynamic information using Location Based Service (LBS) technology. Using Personal Digital Assistants (PDA) and Global Positioning System (GPS), tourists can acquire cultural information when they approach a place of cultural heritage. A GIS database of Deoksugung Palace, one of the most famous areas of cultural heritage in Seoul, South Korea, was constructed and dynamic functions connecting location information with cultural content in the GIS database were implemented. As a result, it is expected that the dynamic system will contribute to tourist industry by providing location-related cultural information to tourists.

1 Introduction

LBS can integrate various mobile hardware devices, wireless communication networks, industry specific software applications with geographic information to provide users with location-related guidance. As the wireless communication and GPS technologies have continuously advanced, LBS expands its application areas to military and government industries, emergency responses, and other commercial sectors (Montoya, 2003; Spinuzzi, 2003; Schiller and Voisard, 2004; Wiafe and Davenhall, 2005). Tourist industry can be also one of commercial application areas that have potential market values for LBS (Schiller and Voisard, 2004); the study on specific LBS applications in tourist industry, however, is a new challenge in academic and industrial communities. In recent years, various GIS-based applications have been developed in the tourist industry (e.g., Akcay and Altan, 2003). They were focused on only static information without consideration of the real-time location of tourists. The real-time location of tourists is one of the critical factors for interactive and dynamic information services; the development of new LBS-based application is needed. This paper addresses a prototype of cultural inheritance information system for more interactive and dynamic tourist information services. The following section presents an example of the system on the Deoksugung Palace in Seoul, South Korea.

2 Motivation and Site Information

Recently, the number of foreign tourists visiting Korea has increased and tourist industry of Korea has expanded its market areas. Comparing from the year of 1996, there was a 56.8 % expansion of tourist industry of Korea in 2004 and a further expansion is expected this year. Despite its obvious growth, however, the quality of information services for foreign tourists has not improved. Only a few inheritances in metropolitan areas provide volunteers and tour guides who are good at foreign language. Moreover, foreign tourists can use the cultural inheritance services at the limited space and within the fixed time schedule. Tourists need more flexible services which can provide useful information about cultural inheritances any time and any where. LBS technology can be an optimal solution to improve the service quality, because it is suited to ubiquitous service framework (any time and any where). Using novel Information Technology (IT) infrastructures of Korea, successful development of LBS-based application for tourist industry can be possible.

In this study, Deoksugung Palace is used for displaying example of cultural inheritance information system. Deoksugung Palace (Fig. 1) was built in 1593 and it is one of the most famous cultural inheritances in Seoul, South Korea. Deoksugung Palace is, among other things, a palace famous for its elegant stone-wall road. It is the only one that has western style buildings beside it, which adds to the uniqueness of the scenery (Kim, 1994).

Fig. 1. Panoramic view of Deoksugung Palace

3 System Architecture

Figure 2 shows the system architecture of the cultural inheritance information system. The system consists of mobile clients running on the Window CE-based Pocket PC and a GIS server running on desktop PC. Mobile clients provide the on-site inheritance information to tourists. A GPS receiver is linked to the Pocket PC through a Bluetooth connection, and is used to capture the location of tourist. Based on the location of tourists, the cultural inheritance system operates in the PDA. As a result, using the PDA and GPS, the on-site inheritance information system provides tourists various information of cultural inheritance.

GIS running on desktop PC is the data management and analysis subsystem which is a decision-making system. This system provides inheritance administrator with tools for managing the application database as well as statistical analysis of tourists' main path and length of their stay.

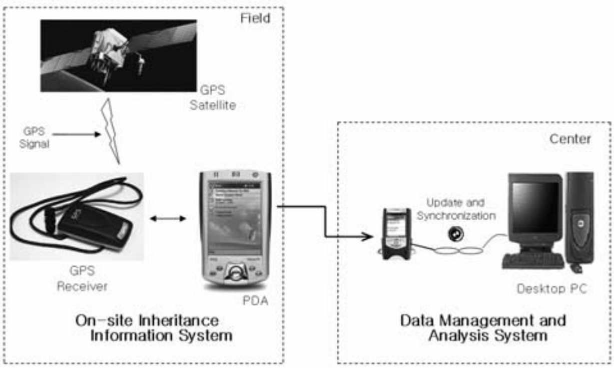


Fig. 2. Architecture of the system

4 System Implementation

4.1 Database design and construction

Database of cultural inheritance information system consists of a digital map and pictures of Deoksugung Palace, documents, and the guidebook of Deoksugung Palace (Fig. 3). Based on these data, five thematic layers of Deoksugung Palace were created and assembled in GIS: (1) pedestrian passage layer, (2) circumference layer, (3) facility layer, (4) Deoksugung Palace layer, (5) main polygon layer. Main polygon layer includes spatial and historical information of cultural inheritances, facilities, and a pedestrian passage. All five layers were created in vector data structure with first two using polyline, and next three using polygon, respectively.

Fig. 3. Construction of GIS database and map

4.2 User interface

Figure 4 shows the user interface of the cultural inheritance information system. The user interface includes toolbars (i.e., main toolbar, browse toolbar, cultural inheritance information toolbar) Tools of cultural inheritance information toolbar, from left to right, are used to show Deoksugung Palace map, display location of the information center and facilities, exhibit recommended tour paths, identify information of cultural inheritance, and start Cultural Inheritance Information System.

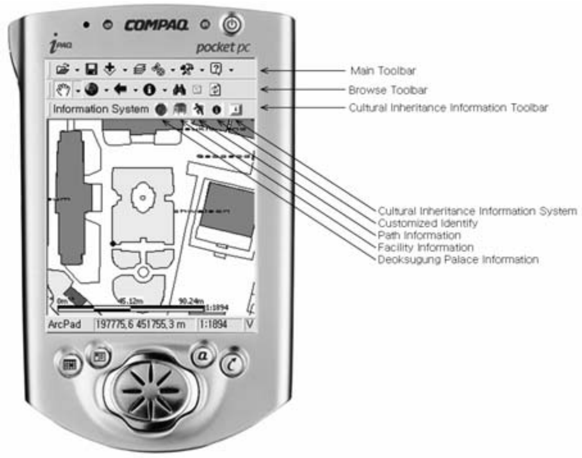


Fig. 4. Customized toolbars of the system

4.3 Main service 1

As shown in Figure 5, clicking the "Deoksugung Palace Information" icon will present layers related with Deoksugung Palace. These layers show information about arrangement of cultural inheritances, facilities, and a pedestrian passage. When these layers zoom in below 1:3000, the labels of inheritances and facilities will be presented.

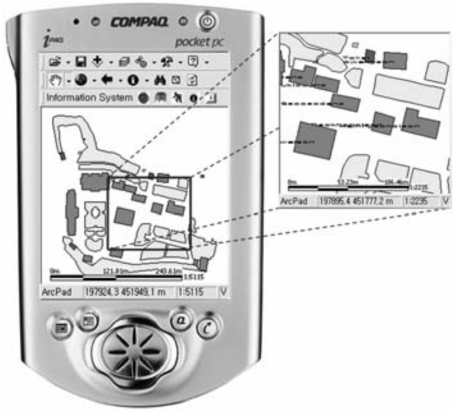


Fig. 5. Interface of the Deoksugung Palace Information

4.4 Main service 2

Figure 6 displays that clicking the “Path Information” icon will present layers related with Deoksugung Palace and a pedestrian passage. These layers show information about location of cultural inheritances and recommended tour paths. The recommended tour paths are the shortest courses which enable to watch all cultural inheritances of Deoksugung Palace. When these layers zoom in below 1:3000, the labels of inheritances and facilities will be presented.

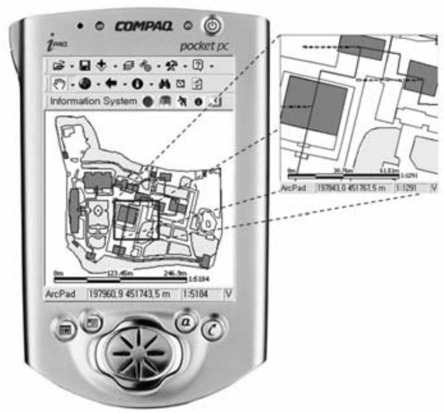


Fig. 6. Interface of the Path Information.

4.5 Main service 3

Figure 7 shows that clicking the “Cultural Inheritance Information System” icon will activate customized inheritance information system. Using ArcPAD Application Builder and the GPS technology, the layers related with Deoksugung Palace and the tracking information of the tourist is displayed. Through the customized system, the tourist can acquire various dynamic information of cultural inheritance when he approaches cultural inheritance below five meters.

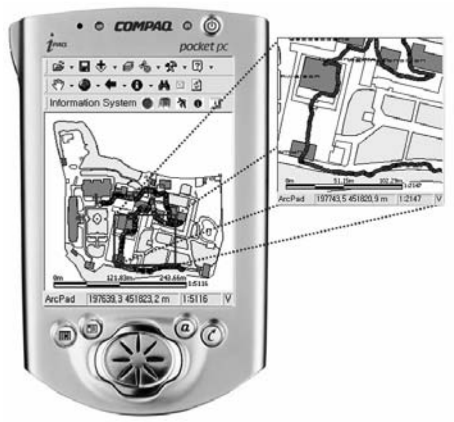


Fig. 7. Interface of the Cultural Inheritance Information System

5 Conclusion

This paper presents a prototype of cultural inheritance information system for more interactive tourist information services. The system consists of two components: the on-site client and desktop server for management and analysis. The system records the location of tourists, and automatically connects the location information with cultural inheritance contents of GIS database. A pilot study on Deoksugung Palace shows that the developed system can provide interactive information with tourists visiting the cultural inheritance and effective tools for facility analysis with managers of cultural inheritance. As a result, it is expected that the dynamic system will contribute to promote tourist industry by providing location related information to tourists.

References

Akcay, O., and Altan, M.O., 2003, Design of Tourist Information System (TIS) and Integration with Panoramic Imaging, Proceedings of 19th CIPA, Antalya, Turkey, pp. 157-160.

Kim, S., 1994, Deoksugung Palace, Dae-Won Press, Seoul, pp. 45-48.

Montoya, L., 2003, Geo-data acquisition through mobile GIS and digital video: an urban disaster management perspective, Environmental Modeling & Software, Vol. 18, pp. 869-876.

Schiller, J., and Voisard, A., 2004, Location-Based Services, Morgan Kaufmann Publishers, San Francisco, California, pp. 10-13.

Spinuzzi, C., 2003, Using a Handheld PC to Collect and Analyze Observational Data, Proceedings of SIGDOC'03, San Francisco, California, pp.73-79.

Wiafe, S., and Davenhall, B., 2005, Extending Disease Surveillance with GIS, ArcUser April-June, ESRI Press, Redlands, California, pp. 14-17.

LWD-Infosystem Tirol – visual information about the current avalanche situation via mobile devices

Michaela Kinberger[1], Karel Kriz[1], Patrick Nairz[2]

[1] Department of Geography and Regional Research, University of Vienna Cartography and GIS, (kinb,kriz)@atlas.gis.univie.ac.at
[2] Avalanche Warning Center Tyrol, Austria, p.nairz@tirol.gv.at

Abstract

The Avalanche Warning Center Tyrol (Innsbruck, Austria) together with the Department of Geography and Regional Research at the University of Vienna have developed a complex and very powerful database driven online decision support system for visualization and analysis of current avalanche relevant factors in the Tyrolean Alps. In order to understand the avalanche situation it is important to have spatial coverage of meteorological and snow pack factors as well as information covering the avalanche danger scale and topographic situation. All information can be interactively made accessible to the user and includes for example current snow depth, amount of snow accumulation within the last 24 hours, temperature, wind speed and direction as well as the regional distribution of the avalanche danger scale including height and temporal dependencies. Spatial depiction of this information can help comprehend the situation. The faster this information is made accessible the more useful it can be. For this reason the partners of the project decided to adapt the online maps for the presentation on mobile devices like mobile phone and personal digital assistant (PDA).

In December 2004 the first images were presented on mobile devices. The main problem was to provide the same information included in the

Internet maps on a significant smaller display. The results are simplified depictions of the current snow depth, amount of snow accumulation and the predicted avalanche danger scale for small screens in two different sizes. This contribution will give an overview of the application, focusing on its latest developments as well as planned extensions.

1 Introduction

A permanent changing environment requires a rapid flow of information to users and decision makers. Modern information technologies like the Internet and mobile communication makes it possible to collect important information and to publish reports about the current state of our environment almost in real-time. By incorporating GIS and cartographic expertise relevant facts with spatial relation can be depicted in an easily understandable visual form. The data transfer via Internet and SMS/MMS to a computer screen or the display of a mobile phone guarantees a fast circulation of current information. For different user devices with a screen size from 120x160 up to 1024x768 pixels the same geodata have to be preprocessed, selected and visualized in a different way. The aim is the production and quick update of maps for visual communication of avalanche relevant facts covering mountainous regions.

2 Topic

One of the main duties of all Avalanche Warning Center's today has not changed since their foundation – prevention of avalanche fatalities by informing the public about the current snow and avalanche situation in different regions. The big difference to former times is however the utilization of new possibilities and techniques of data-collection, spatial depiction and information transfer.

The Avalanche Warning Center Tyrol has developed into a high-tech institution with comparatively high-quality standards. Due to sufficient financial support by the local government there is not only an exhaustive network of observers but also one of the highest densities of high-alpine automatic weather stations in the world (see figure 1).

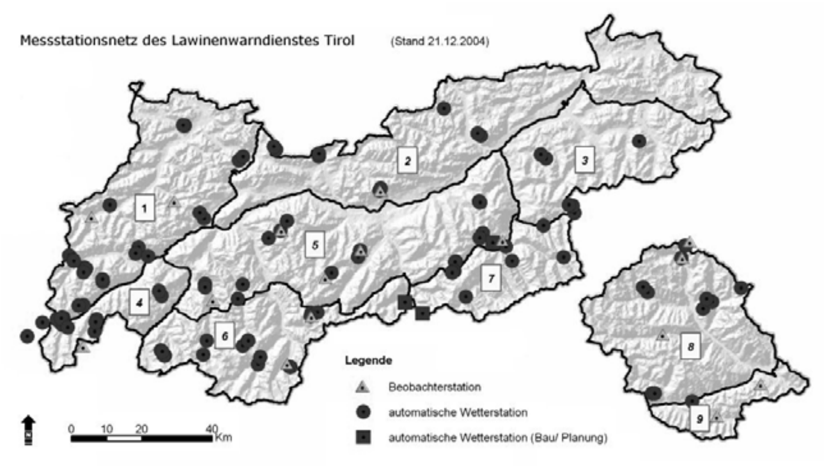


Fig. 1. Network of observers and automated weather stations

Since the first construction of such weather stations a lot of experience has been achieved. Surprisingly enough their reliability in severe conditions is amazing, however their limits become also visible under such circumstances. In order to receive reliable data, which is one of the bases for work and the herein described developments, a comprehensive analysis of every potential location has to be undertaken before building them. The succeeding correction of data even with special software is much too time consuming and imprecise.

Some years ago the Avalanche Warning Center Tyrol started a very fruitful collaboration with the Department of Geography and Regional Research at the University of Vienna. At the beginning work concentrated on verifying and formatting the available meteorological data. The next step was the automatic transfer of confirmed data by the Avalanche Warning Center to the Department of Geography and Regional Research. The Department itself developed an online decision support tool where different types of maps are made accessible during the winter season for any given time period. The aim was a faster and easier perception of important snow- and avalanche-relevant data. The results should assist the forecaster's work as well as support the user's needs and hopefully help reduce avalanche accidents.

In the winter season 2003/2004 the online decision support tool provided cartographic visualization of the current snow depth, the amount of snow accumulation within the last 24, 48 and 72 hours as well as the regional distribution of the avalanche danger scale including height and temporal dependencies. To reach an even broader user group it was necessary to expand the scope of the project and provide a cartographic depiction for the access with mobile devices. After a methodical and conceptual processing phase the first maps for smaller displays were ready in December 2004. The maps were then transferred via multimedia messaging service (MMS) to mobile phones and PDA. The main problem was to provide the same information included in the Internet maps on a significant smaller display. The results are simplified depictions of the current snow depth, amount of snow accumulation and the predicted avalanche danger scale for small screens in two different sizes (mobile phone and PDA).

3 Conception and implementation

The advantages of cartographic representations on the Internet can be seen in up-to-dateness, interactivity, spatial communication as well as efficient and cost-effective dissemination. In order to benefit from such an interactive communication platform for environmental strategic decisions support it is important to make use of automated cartographic procedures. This is possible by utilizing standardized cartographic methods with combined access to permanently available thematic geodata that allow the production of spatial output. Thereby cartographic layout and design aspects play an important role within such a workflow. Geometric data is in most cases static and therefore does not heavily influence the overall work procedure. Thematic data such as weather data, that changes rapidly is considered variable and is controlled by a time-dependent process.

In many cases interactive cartographic visualization with graphic software slows down the workflow. The graphic realization can therefore be the bottleneck within such automated systems. Time plays an important role if up-to-date information must be transported to the user. If design and layout take up too much time cartographic output becomes worthless. In order to accelerate the procedure and keep the output time lag low operation of cartographic realization in batch mode can be utilized. With the help of such automated time-dependent systems short-updating cycles and high efficiency can be achieved.

Working in batch mode enables the system to execute tasks without interaction of a user. With the help of batch mode processing a time-controlled system can achieve very short updating times. The advantages are actuality but also in the long run cost effectiveness, consistent quality and worldwide distribution. However one has to keep in mind that extensive development is needed to set up such a system and maintenance work has to be considered in order to keep the system running.

In order to utilize a fully automated cartographic visualization workflow specific system components are needed. These components consist of a user interface for the World Wide Web, a system interpreter that controls the time-dependent procedural workflow of the system, graphic tools that enable a profound cartographic representation, thematic and geometric data that is either accessed internally or externally and stored in a database management system as well as a GIS that controls the interpolated surface creation.

Another possibility to control a process in batch mode is to start the automated cartographic visualization workflow with the occurrence of a certain event. Contrary to the production of maps of the meteorological factors, where the production process begins at a certain time, the maps of the regional avalanches dangers steps are produced event-controlled. Therefore these maps can be updated immediately after the forecast by an expert of the Avalanche Warning Center in Tyrol.

Another point in the conception is dealing with the graphical design and technical limitations of the maps to produce. Nowadays there are plenty of possibilities for implementing maps on the web. It's possible to use different graphic formats, animations and complex layer structure. Above all increasing bandwidth enables lager file size.

Maps for mobile devices are still subject to technical restrictions. Low bandwidth causes limited file size and small displays give only small area for graphical information. To increase the file size all maps for mobile devices are without hills hading. Now it was possible to use only solid colors to visualize the desired information. Therefore the best graphic format to use is GIF (Graphics Interchange Format). Using up to eight different colors we were able to produce files of about 16 kB file size for mobile phones and about 34 kB file size for PDAs.

4 Results

Besides Internet portals, mobile information services are important for the distribution of short living information. With the adapted Internet maps for mobile device access, the user is able to receive important information by the means of modern, locally independent communication technology. In the following the main focus points of work during the last winter (2004/2005) will be described. More examples can be found on the homepage of the project partners (see References and Links).

4.1 Maps for mobile devices

In the summer of the year 2004 the project team decided that users of mobile devices, such as mobile phones and PDAs, should also get access to the information provided by the "LWD-Online-Maps". Today many mobile devices have a color display, even if it is very small. Therefore it is possible to receive and view a simplified version of the online maps on any mobile device.

However, the maps available must be optimized for the small screens of mobile devices. It is absolutely necessary to reduce the content of the graphics without loosing information, so that the cartographic depiction can be represented on the substantially smaller area. The generated maps were integrated in a mobile information system. Members of the local avalanche committees receive daily updated representations of the current snow depth, amount of snow accumulation and the predicted avalanche danger scale on their mobile phone.

Due to the smaller size of the display important topographic elements used in the Internet maps had to be excluded from the maps for the mobile view. In the map of the predicted avalanche danger scale the user can localize his area of interest by following the region boarders. In the maps of the current snow depth and the amount of snow accumulation within the last 24, 48 and 72 hours the selected hydrographic network makes orientation easier.

Avalanche danger in particular regions is often time dependent. To show the changing avalanche danger scale over time animation is used for the Internet maps. The mobile maps are not animated, instead two single depictions are sent to the user, including one picture with "AM" for the predicted avalanche danger scale in the morning and the other one with "PM" for the predicted avalanche danger scale in the afternoon. All mobile maps are depicted in two different sizes. For the smaller mobile phone display pictures with the size of 133x100 pixel are produced (see figure 2). For PDA displays images with the size of 200x150 pixel are generated.

Fig. 2. Maps for mobile phones

4.2 Maps representing the current air temperature, wind speed and direction

Besides the information on the current snow depth and the accumulation within the last 24 hours the automated weather stations in the Tyrolean Alps also collect values on the current air temperature, wind speed and wind direction. All these meteorological factors are transferred and stored in a centralized database.

For the visualization of the current air temperature in degree Celsius the values of the single weather stations are interpolated dependent on their heights, classified and colored (see figure 3). The class limits depend on significant values of the air temperature that are important for the formation of the snow pack. The class borders can still be changed at a later time by the request of experts.

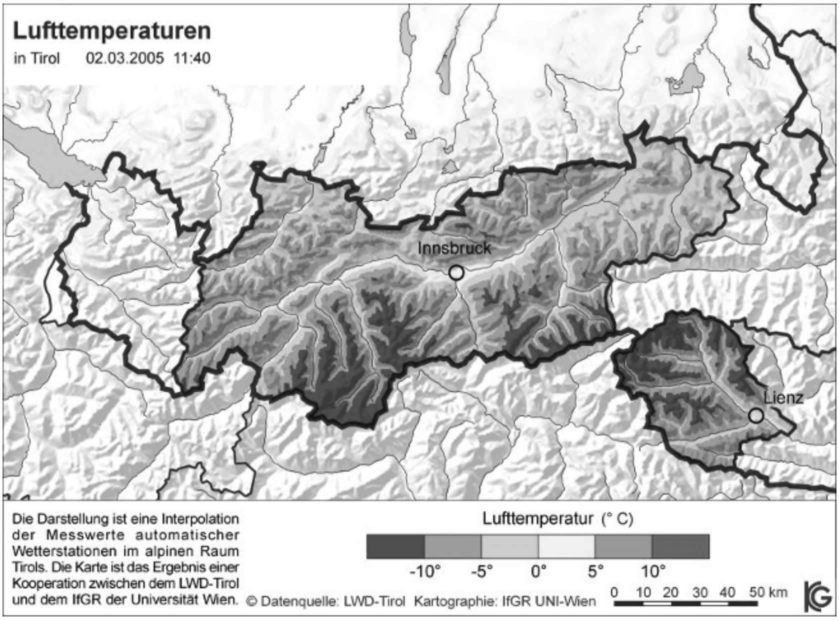


Fig. 3. Map representing the air temperature

Current wind speed and wind direction are represented in one map. An arrow signature, pointing at the geographic position of the weather station, gives information about the dominating wind direction. The wind speed is depicted by the color of the arrow.

4.3 Maps representing the regional avalanche danger scale for color-blind people

In order to assist color-blind users and in agreement with experts, the necessity of a grayscale cartographic representation of the regional avalanche danger scale was implemented (see figure 4). Up to now these maps for color-blind users were only generated for the Internet, but not for the access with mobile devices. A variation of the grayscale representation for smaller displays is considered for the coming season.

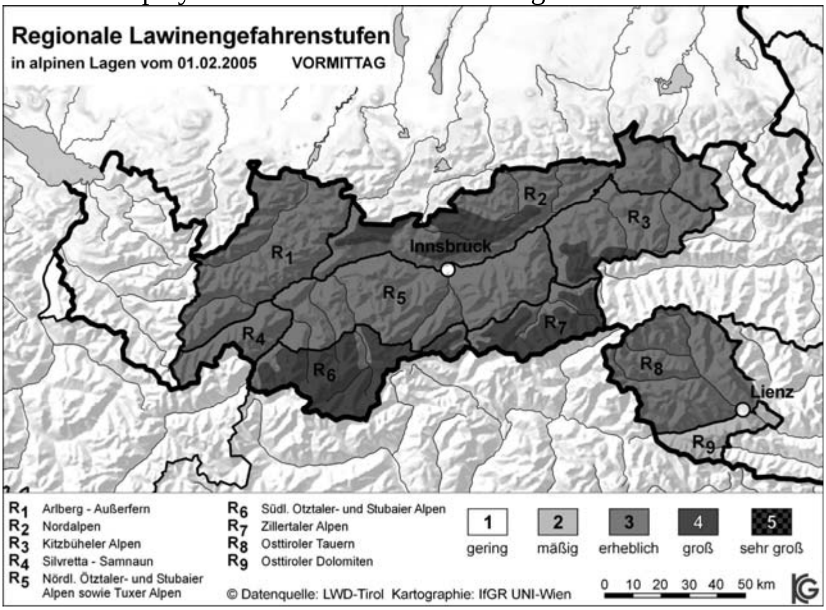


Fig. 4. Grayscale representation of the regional avalanche danger scale

5 Outlook

A permanent changing environment requires a rapid flow of information to users and decision makers. This paper describes one possibility of prompt and efficient cartographic visualization utilizing the web and mobile information services that can be accessed for environmental issues such as avalanche forecasting. To deal with the technical limitations of mobile devices, map elements depicted with graduated tints (e.g. hill shading) were omitted. Only a few solid colors were used to visualize the desired information. The produced maps are GIF graphics of small file size.

During the past winter seasons the reliability and the accuracy of the online maps produced have been monitored. The high amount of user access showed the tremendous public interest in online visualization of the snow and avalanche danger situation in Tyrol. In the future it should be possible to provide this important information not only for Tyrol but for the entire Austrian Alps.

References and Links

Kinberger, M., Kriz, K, Nairz, P. (2003): Automationsgestützte kartographische Visualisierung im Internet - Ein Hilfsmittel für den Lawinenwarndienst. Strobl/ Blaschke/Griesebner (Hrsg.): Angewandte Geographische Informationsverarbeitung XV. Wichmann Verlag, Heidelberg, pp. 218-223.

Kinberger, M. (2003): Datenbankgenerierte Internetkartographie – Ein Hilfsmittel für den Lawinenwarndienst. Kartographische Nachrichten, Heft 5, Kirschbaumverlag, Bonn, pp. 203-206.

Kriz, K., Nairz, P., Kinberger, M. (2004): LWD-Infosystem Tirol concept and design of an avalanche decision support system. Proceedings, 4th ICA Mountain Cartography Workshop, Vall de Núria, 30th September - 2nd October 2004, Barcelona, pp. 1-12.

LWD Online Maps: http://www.gis.univie.ac.at/lawine

Tiroler Lawinenwarndienst: http://www.lawine.at/tirol

Spatial tracking in sport

László Zentai[1], Antal Guszlev[2]

[1] Department of Cartography and Geoinformatics,
Eötvös Loránd University, laszlo.zentai@elte.hu
[2] Faculty of Geoinformatics, University of West Hungary
a.guszlev@geo.info.hu

Abstract

This chapter deals with the possibilities of using Location Based Services (LBS) in different kinds of sports. The use of LBS enables movements to be tracked easily and objectively. This advancement in technology literally brings new dimensions to popular outdoor and indoor sporting activities. In the following we describe how these new ways of positioning have affected the sports of soccer, orienteering, rally, ballooning and sailing.

1 Overview

The history of sport is probably as old as the existence of people as purposive beings. Sport has been a useful way for people to increase their mastery of nature and the environment. It seems to involve basic human skills being developed and exercised for their own sake, in parallel with being exercised for their usefulness.

Sporting events are organised not only for the competitors, but also for the cheering crowd. They captivate the minds of the general public and sports are followed by millions of people around the world: in front of television screens, internet browsers, but also PDAs and other handheld devices. The German mobile communication providers await the Football

World Cup 2006 with great expectations: it should give a thrust to the 3G boom as hoards of mobile phone users will be capturing and sharing the most memorable moments and the biggest goals.

In order to attract many viewers, it is important to focus on the requirements of the people and the media. Some traditional sports even change their rules to make it more visible, more exciting for the audience and/or for the televiewers. Sport is more and more a business and staying attractive and interesting is a vital challenge for most of them.

Citius, Altius, Fortius! – Faster, Higher, Stronger! The ever growing demand stated in the Olympic motto suggests that 3D position, time and velocity are key factors in precise and objective sport evaluation. In the past years we got used to technological advancements in the sports science industry. Basketball and figure skating broadcasts show video replays of major accomplishments. Tennis has the net sensor and the automatic eye on the line. Ice hockey has the goal sensor and athletics have the electronic starting blocks and finish line. It is important to note however that because of the high cost and complexity, these innovations have been introduced at the highest levels only.

2 Rally

Rally racing is a form of automobile racing running not on a circuit, but instead in a point-to-point format where participants and their co-drivers “rally” to a set of points, leaving in regular intervals from start points. The entertaining and unpredictable nature of the stages draws massive spectator interest.

To aid the navigation of the racers, calibrated odometers with large displays are generally used inside the cars, and can be preset quickly to give a countdown of the distance to the next feature. Most rally cars are also equipped with GPS receivers for a number of years now. The high speed and the intricate terrain enable a general accuracy of around 10 metres. These devices are used not only for tracking, but also to improve safety.

However, if we want to render an account of lost seconds, and split-seconds, we have to reconstruct the track in an extremely high detail. In 2002 the FIA World Rally Championships (WRC) teamed up with Inmarsat, who became the championship’s first exclusive global partner. In the German round of the 2005 (WRC), competitors tested the accuracy of the European Geostationary Navigation Overlay Service (EGNOS). By storing and post-processing the improved EGNOS positional data, the whole stage had been reconstructed and played back in a 3D virtual environment. The accuracy of the map and the terrain model was higher than most commercially available road maps which only show the center vertices. The test proved positive; EGNOS signals were reliable and consistent, providing the necessary corrections. Using techniques originally developed for the America's Cup in New Zealand and a mix of off-the-shelf and custom software, Virtual Spectator recreates portions of the rally from the GPS logs. The video game-style animations can combine the progress of multiple cars to give broadcast viewers a head-to-head comparison.

Fig. 1. A 3D display showing positions derived from EGNOS and standard GPS (Credits: ESYS, MQL, Inmarsat)

Current GPS systems are unlikely to replace the role of the co-pilot and his 'pace notes', because the delicate cooperation between the driver and the co-pilot can hardly be automatised. The co-driver has to read the notes at such a pace that the driver will always have enough to go on, but not so far ahead that he will have forgotten it by the time the feature is encountered.

3 Orienteering

Orienteering is a cross-country sport involving navigation with a map and compass. The runners attempt to visit, in sequence, control points that are indicated on the map, as quickly as possible. One of the key elements of orienteering sport is the route choice: the shortest way in time between to control points depends on different factors (weather, training, physical condition of the competitor). Orienteering has never been very popular in most countries (except Scandinavia) and it is not present in television. The orienteering sport has some specialities, which do not make the media friendly transformation easy:

- Orienteering events mostly take place in remote, forest-covered areas, where to assure the broadcasting infrastructure is not easy or expensive.
- The competition map and course is secret. If we give an on-line broadcast we have to have extra effort to avoid unwanted information handover to the competitors. The preparation work of the broadcast may require some outsiders to be included in the secret part of the organizing.
- Orienteering maps are very special products; they are interesting and legible only for the “experts” (competitors, fans): we have to transform the orienteering map to a media friendly, attractive screen image to make the broadcast interesting. This is a real TeleCartography task: how can we visualize the reality for the viewers.
- In order to follow the runners’ route, the competitors have to carry a GPS receiver with them and another (GSM/GPRS) device to report their position. These devices must be very small and they cannot affect the competitors’ performance.

Typical problems in route tracing:

- How can we get GPS signals on steep slopes in the forest, where the receiver cannot see enough satellites?
- How can we send back the competitor’s position if there is no reliable GSM coverage in the area?

The basic operation of the system designed for the 2001 World Orienteering Championships (WOC) in Finland was:

- The Benefon Track GSM received signal from GPS satellites and determined the position of the competitor every second.
- The position was sent in every 20 seconds as an SMS via the special network.
- A server connected to the short message centre received the position and stored it in a special database.
- According to the instructions from the TV director, predefined parts of the map with selected competitor(s) were retrieved and their route drawn with special software as a still image including the competitor names.

- The image was transferred into the system to be seen by the TV director and being available for live broadcasting 15 seconds behind the real situation (nearly real-time) on the average.

The GSM network was configured for the competition area so that the signalling capacity was maximized to ensure delivery of SMS's. The signalling load of tracking SMS traffic was monitored during the tests and it was concluded to be adequate for the WOC events. Sonera (a Finnish telecom provider) had a local base station in the event centre during the WOC 2001 events so that the GSM traffic by the audience did not give any load to the competition area.

In general, the system worked the way planned and tested and the project goal was considered achieved.

- The amount of captured tracking routes was less than expected. The main reason for this was the GPS geometry during the competition. This had some impact on the TV visibility, since 2-3 images requested by the TV director were discarded due to incomplete or missing data. It has been concluded that nothing could have been done to improve this. In this geographical area, in a Finnish forest with constantly moving objects, the GPS geometry is the key factor that cannot be compensated within the GPS device. GPS geometry can be estimated beforehand and for better results the race should have taken place in an other time. It is impossible to estimate the GPS geometry accurately more than 6 months in advance.
- The TV picture generation in the software became slower when the amount of data increased in the systems database due to a software design error. This error also caused a decrease in the TV visibility, since the timing is the key factor in a live TV broadcasting. The TV setup was larger than ever used before in orienteering.
- The used technology led to a system that was not as much a real-time system as the TV director had wished. Using SMS technology and a relative slow update frequency is best suited to see the overall picture of route choices and mistakes, but is not suitable for providing data into a really live broadcasting that tries to capture the struggle and live action.
- The tracking system did not show any dramatic mistakes of competitors in TV. The main reason for this was not a bad working system, but rather that there were very few mistakes and different route choices made by the top runners.

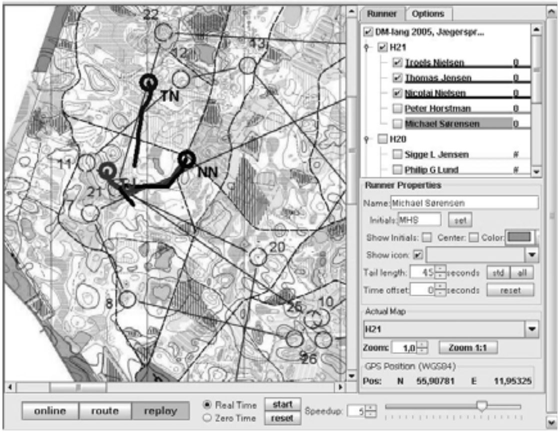


Fig. 2. The TracTrac-applet is a software that allows you to see the route and position of the participants of sport competitions and events where the TracTrac system is used. The runners carry a combined GPS/GSM-unit on the back. The device is very light (117 grams) and is worn in a small strap on the back. Due to the light weight and flexible strap, the unit can be worn without compromising performance.

3.1 Displaying the orienteering map on the screen

We have to deal with the screen representation of orienteering maps to let the less trained spectators, televiewers enjoy (or even just understand) the broadcast. There are too many details in orienteering maps: the spectators cannot interpret all details, so the transformed map must show clear and simplified alternatives. The three dimensional representation can make the image more interesting, but it is more difficult to generate and it can be more complicated to interpret such representations for the audience. We have to find a good compromise.

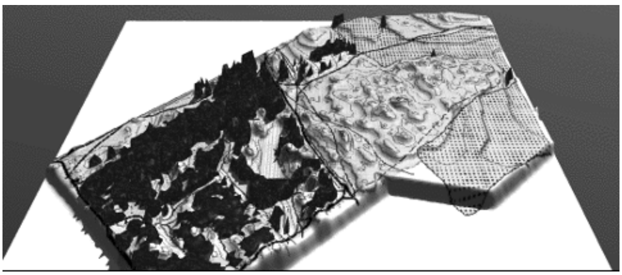


Fig. 3. 3D image generated from an orienteering map (courtesy of Tamás Heckenast, Hungary)

The 3D terrain representation must be simplified and it is difficult to show the whole area on a screen. It is necessary to create different maps:

- a most simplified one to help the general overview of the area (we can add virtual landscape features like trees, buildings, but it is not suitable to show the course);
- less simplified ones to show only a part of a course (one-two legs with different route choices).

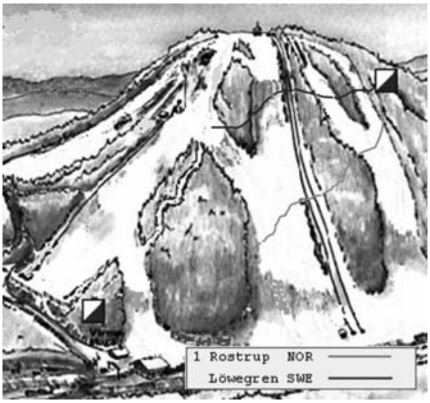


Fig. 4. 3D-like representation of an orienteering map showing route choice

Orienteering is regularly shown in television broadcast in the Nordic countries, so they have enough experience how to do it efficiently. In most cases there is no time for map preparation or it would require too much extra effort or the organizers are keen on secrecy (and very few orienteering events are so important that they get remarkable broadcast time, which would give the opportunity to present competition maps). In exceptional cases tracking system with minimal map preparation were used. In the Nordic Championships (2005, Norway) a new, special form of orienteering were introduced at international level (micro sprint). This new form was unknown for the audience so a graphic representation helped the audience to understand the new rules.

4 Soccer

Football or soccer is a team sport played between two teams of eleven players each. The objective of the game is to score by manoeuvring the ball into the opposing goal. Football is played at a professional level all over the world, and millions of people regularly go to matches to follow their favourite team, whilst billions more avidly watch the game on television. A very large number of people also play football at an amateur level.

The need for providing the soccer referee with accurate and quick information to assist his administration is not a new one. The imperfections of the man in charge have been under criticism since the game began but with introduction of television and the World Cup Final of 1966, which provided a most graphic example of how the mistakes of one man can alter the outcome of a game, we now enjoy an amount of analysis and record that prior generations have not. Refereeing mistakes are kept alive by memory like the post war game between Glasgow Rangers and the visiting Moscow Dynamo side who played with twelve men for part of the match. The modern ref has his mistakes preserved for ever for all to see – and brought out whenever there is a major blunder by another.

Fig. 5. Soccer game as a potential area of LBS-application

Team games are complicated activities, which involve much interaction between players. Automated multiple object tracking in large, congested, rapidly changing, and frequently occluded domains as a soccer pitch is a complex problem, particularly given the non-linear nature of each player's movements. Several companies began investigating options, including using lasers and light barriers to determine the ball's and the players' position.

Experiments are taking place to put microchips inside the ball and the shoes or leggings of players. For example the chip designed by Cairos Systems weighs only 12 grams with batteries, transmits 2,000 signals per second to a receiver network of 12 antennas, placed around the pitch, including on light fixtures. The receivers then send information about the ball's location to a central computer in real time. After processing the position, 3D coordinates can indicate when the ball crosses the goal line in mid-air.

Micro-chipped players and football tracked by radio signals can assist the referee in near-real time in the following tasks:

- setting the time of the match by recording the duration of the ball in play;
- setting the distance of the defensive wall from a free kick position;
- monitoring the movement of the goalkeeper and taker at penalty kicks
- noting the departure of the ball from the field of play for throw-ins corners or bye kicks;
- keeping the administration of the match simple;
- giving information on the movement of players in an offside situation;
- recording the all important goals.

Apart from the referee, the trainers and broadcasters can find positioning microchips a helping aid:

- trainers can analyse the performance of players, see their weak and strong points and develop technical and tactical trainings.
- broadcasters' cameras are able to follow the players and the ball automatically, and are able to reconstruct and demonstrate 3D animations of important moments.

5 Ballooning

Hot air balloons are the oldest successful human flight technology, dating back to the Montgolfière brothers' invention in France in 1783. In competition, the pilots need to be able to read different wind directions at different altitudes. More experienced pilots are able to take a flight in one direction, rise to a different altitude to catch wind in a returning direction. With experience, luck, and the right conditions, some pilots are able to control a precision landing at the destination. On rare occasions, they may be able to return to the launch site at the end of the flight. This is sometimes called a 'box effect', usually when flying in valleys with drainage winds.

During a race, competitors can have up to 20 different tasks to complete. These include navigating to predetermined positions (e.g. above a road crossing) and dropping special marker ballasts. These markers are then collected on the ground and the scores are inversely proportional to the distance between the marker and the goal.

Fig. 6. Traditional maps and flight logs used in a hot air balloon (http://www.europeanballoon.be/en/GB_2005.htm)

The biggest advantage of using GPS in hot air ballooning is that on-board observers will be completely unnecessary. Loggers provide a more objective and cheaper solution for evaluating the flight. The first official use of GPS was in 2000, at the Gordon Bennett race in Belgium. This was the time when the ’Selective Availability’, the intentional degradation of GPS signals by the US government stopped. The track logging worked fine except for one competitor where the track stopped halfway during the flight. The introduction of the loggers brought unprecedented changes to the air ballooning society. Some conservative racers are very pessimistic and fear that these changes will spoil the traditions and on the long run, it may destroy the sport itself. Though the support for involving new technology is apparently stronger than the conservative view, as it has more advantages then problems.

According to the new rules provided by the Federation Aeronautique Internationale at no time is the competitor allowed to open or interfere with the logger or its operation. The competitor is responsible for any loss or damage between handing over and return of the logger. In case of unusable track logs, the officials may ask the competitor to provide any GPS equipment he may have to substitute the missing track information. It is therefore in the competitor’s interest to equip himself with a GPS that provides track information usable for scoring (position, altitude and time).

The GPS logs the position with an average of 10 metres accuracy. Some argue that the logger is hanging about 5 m from the centre of the balloon and about 7 m above the bottom of the basket, but experiences show that these differences can be neglected. So instead of the accuracy, reliability of the GPS caused the most concern to the pilots (and the scorers). Flying with two devices eliminated most of the problems but some concerns still exist over the recurrence of human error setting these up and the GPS keeping the settings. The reliance on GPS leads to a dependence on IT for success. A network is necessary to achieve the through put of loggers and storage of the data.

An interesting experiment has taken place at the Balloon Trophy 2005 in Murau, Austria. A special application has been developed to track the balloon race in real time. The GPS enabled mobile phones sent the coordinates to a central database server through a GSM network. The actual position and data about the balloons were displayed in a Web-GIS application using an interactive SVG interface. To indicate the height differences between the competitors, a special cartographic solution was applied. 3D positions were mapped in bird's eye view – the size of the image representing a balloon was proportionate to its altitude.

Fig. 7. Example of Live balloon tracking system: Balloon Trophy 2005 in Murau. (http://geoweb02.cti.ac.at/STS_Balloon/index.htm)

6 Sailing

Sailing is the skillful art of controlling the motion of a sailing ship across the sea, using wind as the source of power. Back in the 15^{th} century, the development of sailing and navigation techniques played important roles in the great voyages of explorations. Crossing unknown waters and discovering new lands in those conditions required exceptional talent and courage.

Today's round-the-world sailing races are still extremely tough for competitors, they are tests of individual endurance. Ships are equipped with modern navigation and communication devices, and sailors report their positions constantly. Hoards of internet and TV viewers eagerly try to follow the progress for several weeks or months. It is the task of cartographers to visualize the route in a proper way. Below we can see how the choice of projections and thematic mapping methods can make a difference in the perception of the whole race.

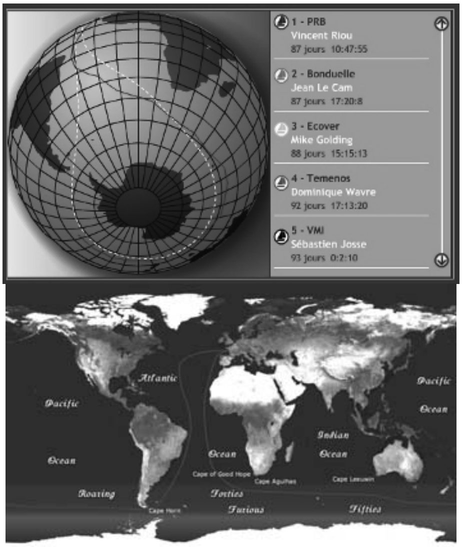


Fig. 8. Two maps of the same route: Vendée Globe yacht race (http://vendee-globe.vendee.fr/parcours/default.asp and http://commons.wikimedia.org/wiki/Image:VendeeGlobeRaceRoute.png)

7 Conclusion

In this paper we selected some sport branches on ground, water and air to give an overview on recent developments in spatial tracking methods. The new technical possibilities are making an enourmous impact on our lives and activities, to which even sports with long traditions are unable to resist. Besides the excitement of live tracking, LBS can serve other purposes also, like bringing objectivity and accuracy for evaluation or enabling further spatial analysis.

The position captured by a GPS device is only the first step. We have to put these constantly changing positions into a comprehensible spatial context, and then broadcast it in a user-friendly way. To make these systems work smoothly and efficiently requires a solid infrastructure, a lot of effort and cooperation: a high level of expertise in IT, telecommunications, GIS and the media is inevitable. The success of these systems heavily rely on the end result: does it fulfil the ever growing needs of participants and viewers?

References

S. Eerola – T. Haanpää: Tracking system – End report Manuscript, Tampere, Finland, WOC 2001

C. J. Needham: Tracking and modelling of team game interactions, The University of Leeds, 2003

K. Dost: Going the Distance: a GPS solution for TSD rallies, GPS World, 2004

FAI Ballooning Commission: AX Model Event Rules for GPS Loggers, 2005

L. Muir: Observer-Less Competitions – A Giant Leap Forward Or A Retrograde Step? CIA Newsletter, Autumn 2005

C. Menard: Web Tracking – GIS basierte Echtzeitverfolgung, FH Technikum Kärnten, 2005

ESA: Satellite navigation tracks rally cars, http://www.esa.int/esaCP/SEMZV9A5QCE_index_0.html, 2005

Links

http://en.wikipedia.org (History of sport, Rally, Sailing, Orienteering, …)
http://www.tractrac.com
http://www.cairos.com
http://www.motram.com
http://www.softsport.com
http://www.debruijn.de/