2014 IEEE 16th Conference on Business Informatics

# Towards Big Data Visualization for Augmented Reality

Ekaterina Olshannikova[1,2]
[1]United Print Center LTD,
Novgorodskaya str. 13, 191144,
St.-Petersburg, Russian Federation
olshannikova.ekaterina@gmail.com

Aleksandr Ometov[2], Yevgeni Koucheryavy[2]
[2]Department of Electronics and
Communications Engineering,
Tampere University of Technology,
Korkeakoulunkatu 1, FI-33720,
Tampere, Finland
aleksandr.ometov@tut.fi, yk@cs.tut.fi

***Abstract*—This article attempts to summarize challenges in visualization methods for existing Big Data and beyond. Moreover, it classifies types of data, analytical methods, visualization techniques and tools, known up to date, with a particular emphasis placed on surveying evolution of visualization techniques over the past years. Following the identified shortcomings of the existing techniques, the role of human eyes physical limitations on perception of the large amounts of information is evaluated and presented along with discussion on Augmented Realty technology and its capabilities for visualization of Big Data and beyond. Based on the results, a novel approach is proposed – it allows one to obtain represented information segment in a short period of time without any significant data losses by placing the most essential data in the central area of the human eye visual field. This article also discusses impact of new technologies, such as Augmented Reality displays and helmets, on the Big Data visualization.**

## I. Introduction

In a world of Big Data it is extremely important to visualize information in such a way that maximizes perception efficiency. It can be assumed that the term Big Data simply refers to the management and the analysis of large amounts of information. The problem is not that organizations are generating enormous amounts of data, but the fact that most of it is presented in a format consistent with the traditional poorly structured database format: weblogs, text documents, or machine code, such as geospatial data [1]. However, it is stored in a multitude repository. Moreover, the use of the cloud storage or the data centers services is also widely common [2]. Furthermore, companies have the necessary tools to establish the relationship between data in addition to the process of making the basis for meaningful conclusions. Besides the fact that data processing rates are growing all the time, it may appear a situation when traditional analytical methods of information would not be able to keep up. Especially, with growing amount of constantly updated data, which ultimately paves the way for big data technologies [3].

There exist various types of data to which certain tools are useful for the analysis. In particular, in recent years a popular visualization method has developed rapidly for the representation of already processed data. It was a step away from the planar images in the direction of multi-dimensional volumetric visualization. However, the big data visualization evolution can not be regarded as a finished product, inasmuch as new techniques generate new research issues and its solutions.

Current activities take its place to create tools for a person to produce quick and effective results working with large amounts of data. Moreover, it would be able to assess the analysis of the visualized information from all the angles in novel scaling ways. All the above leads us to the fact that future visualization tools would be connected to augmented reality, which will be intuitively suitable for limited perception capabilities.

The following text is organized as follows. In the third section, we describe currently used data processing methodologies and main issues connected to human perception. Next in fourth and fifth sections, modern visualization techniques for massive amounts of information are observed. Finally, its evolution is presented and supported by the proposed oncoming Big Data visualization extension for augmented reality which can solve actual perception challenges.

## II. Big Data Processing Methods

In fact the concept of big data implies working with the massive volume information and varied composition, very frequently updated and relocated in different sources in order to increase efficiency, create new products and improve competitiveness.

Large data sources in the world today are ubiquitous. Data used for processing may be obtained from measuring devices, events from radio frequency identifiers, message flows of social networks, meteorological data, remote sensing, data streams on the location of mobile subscribers, devices, audio and video recording. So, as big data is widely used all over the world a novel and essential research field is established. Actually, the mass distribution of the technology and innovative models using different kinds of devices and Internet services was the starting point for the penetration of big data almost in all areas of human activity, primarily, commercial sector and public administration in addition to the currently developing research vector in this field of studies [4].

Currently, there exist many different techniques for the data sets analysis, mainly based on tools borrowed from statistics and computer science. Today to analyze large amounts of data the most advanced techniques are used: artificial neural networks, models based on the principle of the organization and functioning of biological neural networks; methods of predictive analysis; statistics; Natural Language Processing;

978-1-4799-5779-8/14 $31.00 © 2014 IEEE
DOI 10.1109/CBI.2014.42

etc. The most popular approaches in various industries are listed below [5]:

1) *A/B testing.* When control sample is compared to the other alternately. So the identification possibility arises as the best combination of performance to achieve, for example, the best response of consumers in a marketing proposal. Big data allows a huge number of iterations and thus obtain a statistically significant result.
2) *Association rule learning.* Set of techniques for identifying relationships, i.e. association rules between variable quantities in big datasets. It is currently used in data mining.
3) *Cluster analysis.* Statistical method for the classification of objects into groups with the use of the identification in advance.
4) *Crowdsourcing.* Multiple sources based on the collecting data method.
5) *Machine learning.* Direction in computer science (artificial intelligence), which aims to create algorithms based on self-analysis of empirical data.
6) *Network analysis.* A set of techniques used to set links between nodes in networks analysis. In relation to social networks allows us to analyze the relationship between individual users, companies, communities, etc.
7) *Data fusion and data integration.* It allows to analyze comments of social networks users and to compare the results in real time.
8) *Data mining.* This technique allows defining a category of consumers which are the most susceptible to the promoted product or service. Especially, this methodology helps to identify the most successful employees and predict consumer behavior model.
9) *Optimization.* the set of numerical methods for the complex systems redesign and improvement processes for one or several indicators.
10) *Visualization.* Graphical representation method to visualize the innumerous amount of the analytical results as diagrams or animated figures to make users interpretation gentler.

Currently, big data has been analyzed and processed continuously. Generally, existing methods of the analysis are focused on the processing and storing large amounts of information facilities in addition to the necessary links and patterns identification. The existence of big data leads to certain challenges and problems, including storing and processing data, its availability and information exchange, complexity and unstructured data. The main reasons causing difficulty in processing and analyzing big data are narrow human perception and limitations from the display size in addition to its resolution. The main big data interaction problem is to extract useful portion of the information from its massive volumes. To solve this issue, visualization technique using certain operations is suitable because it transforms the information into accessible and acceptable for human perception one. Nowadays, different groups of people including designers, software developers and scientists are in a process of search for new visualization tools and opportunities. It could be called a new revolutionary movement that allows ones to make the sightless visible as well as to modify existing ideas about the complex world of data.

## III. VISUALIZATION METHODS

There is a fairly large amount of data visualization tools that offer different possibilities. It can be classified as a character of data to be visualized by the tool; visualization techniques and the samples as the data can be submitted; interoperability with visual imagery and techniques for better data analysis.

Visualization tools are able to work with:

- *univariate data* – one dimensional arrays, time series, etc.;
- *two-dimensional data* – point two-dimensional graphs, geographical coordinates, etc.;
- *multidimensional data* – financial indicators, results of experiments, etc;
- *texts and hypertexts* – newspaper articles, web documents, etc.;
- *hierarchical and links* – the structure subordination in the organization, e-mails of people, documents and hyperlinks, etc.;
- *algorithms and programs* – information flows, debug operations, etc.

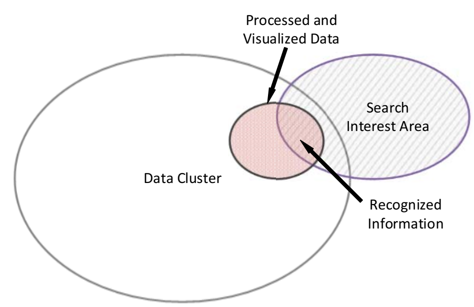


Fig. 1. Human Perception Capability Issue.

However, different visualization methods are currently used to represent various types of data [6]. Obviously, the number of forms is only limited by human imagination. The main requirement is clarity and ease of the analysis of the represented data. Visualization techniques can be both elementary (line graphs, charts, bar charts, etc.) and complex (based on the mathematical apparatus). Furthermore, the visualization can be used as a combination of various methods. However, visualized representation of data is abstract and extremely limited by ones perception capabilities and requests (see Figure 1). Types of visualization techniques are classified as following:

- *2D/3D standard figure* [7] – bars, line graphs, etc. The main drawback of this type is the complexity of the acceptable visualization for complicated data structures;
- *Geometric transformations*[8] – scatter diagram of data, parallel coordinates, etc. This type is aimed to the multi-dimensional data sets transformation in order to

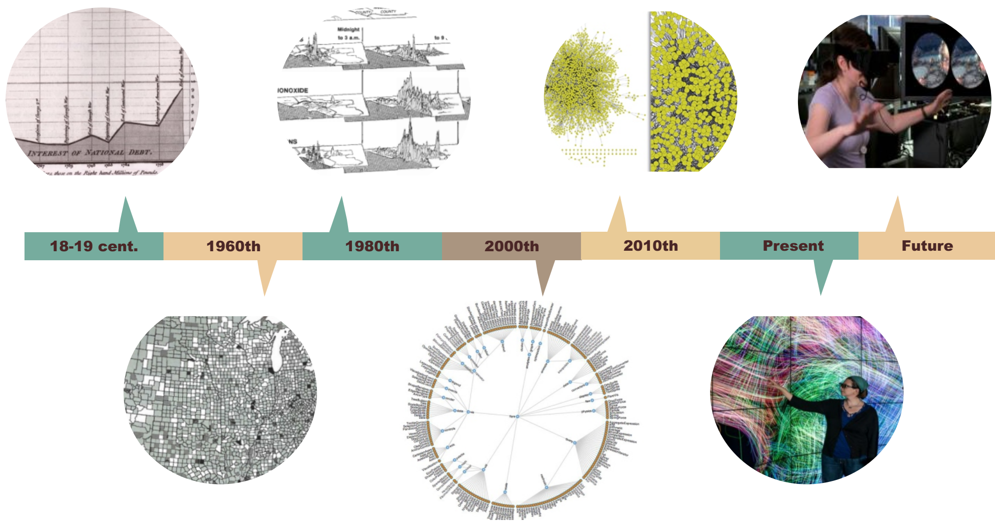


Fig. 2. The Evolution of Visualization Methodology.

display it in the Cartesian and non-Cartesian geometric spaces. This class includes methods of mathematical statistics unit;

- *Display icons* [9] – ruled shapes (needle icons) and star (star icons). Basically, this type is displaying the values of elements of multidimensional data in properties of images. Such images may be: human faces, arrows, stars, etc. Images can be grouped together for a holistic analysis. The result of the visualization is a texture pattern, which is different corresponding to specific characteristics of data;
- *Methods focused on the pixels* [10] – recursive templates and cyclic segments. The main idea is to display the values in each dimension into the colored pixel and to merge some of them according to specific measurements. Since one pixel is used to display a single value, therefore visualization of large amounts of data (over one million values) can be reachable with this methodology;
- *Hierarchical images* [11] – tree maps and overlay measurements. These type methods are intended to be used with the hierarchical structured data.

Visualization methods application should present visual forms that capture the essence of data itself. However, this is not always enough for a complete analysis. Data representation should be constructed in order to allow user to have different visual point of view. Thus, the appropriate interoperability should be performed:

- *Dynamic projection* [12]. Non static change of projections in the research of multidimensional data sets is used. An example of the projection in two-dimensional plane of multidimensional data in a scatter plots. Necessary to note, that the number of possible projections exponentially increases with the number of measurements. Therefore, with a large number of measurements, projection would become harder to perceive;
- *Interactive filtering* [13]. In the investigation of large amount of data there is a need in sharing data sets and highlighting significant subsets – filtering images. It is significant, that the opportunity should be provided in real time with visual representation. A subset can be chosen either directly from the list or by determining a subset of the properties of interest;
- *Scaling images* [14]. Scaling is a well-known method of interaction used in many applications. Especially for big data processing, this method is considerable due to the representation of data in a compressed form volume and it provides the ability to display any part of it in a more detailed form simultaneously. Nevertheless, lower level entity may be represented by a pixel at a higher level, a certain visual image and the following text label;
- *Interactive distortion* [15] supports the research process data using distortion scale with partial detail. The basic idea of this method is that a part of the fine granularity displayed data is shown in addition to one with low level of details. The most popular methods are hyperbolic and spherical distortion;
- *Interactive combination* is bringing together a combination of different visualization techniques to overcome deficiencies of one of it. For example, different points of the dispersion projection can be combined with the techniques of coloring and arranging dots in all projections.

Any visualization method can be classified in all three parameters, thus by type of processing data, visual images that it can provide, and the ability to interact with visual images. Obviously, one method can support different types of data, various images and varied ways to interact with it.

A visual representation of the big data analysis is crucial for its interpretation. Moreover, it is obvious that human perception is limited. So, scientists continue their research in this field with the main purpose of the improvement of modern methods of data presentation in the form of images, diagrams or animation.

Nevertheless, a progressive visualization methods were invented rapidly [5]:

1) *The tag cloud* in the case of text analysis weighting value depends on the frequency of use (citation) of a particular word or phrase.
2) *The klastergramma* is an imaging technique used in the cluster analysis. Shows how individual elements of the data relate to the clusters as they change their number. Choosing the optimal number of clusters is also an important component of the cluster analysis.
3) *The spatial flux* is a chart which allows one to monitor the spatial distribution of information.

In addition, classification of the information processing stages is listed below [16]:

1) visualization of the source data;
2) purification of the data and to preliminary treatment;
3) control of processes by means analysis of visual interactions;
4) results of analysis by interactive visualizations.

Using visual and automated methods in processing big data allows ones to use human knowledge and intuition in processing extensive amounts of information at all stages. Moreover, it becomes available to reach the novel solutions discovery for big and complex data [17]. Vast amounts of information motivate researchers and developers to create modern tools for quick and accurate analysis. As an example, the rapid development of visualization techniques may be concerned. However, in current world of the interconnected research areas one needs to combine basic and effective opportunities existing visualization methods, as well as the new technology opportunities to solve main problems and challenges of big data analysis.

## IV. Evolution of the Visualization Methods

Visualization methods evolved a lot over the decades (see Figure 2). During most of this process, the only limit for novel techniques was human imagination. The first methods were performed as a simple plots in the $XVIII$ century followed by one dimensional histograms [18]. Still, those examples are useful only for a small amount of data. By introducing more information, this type of diagram would rich a point of the worthless. However, next step which was presenting a system with additional time dimension appeared as a significant breakthrough. Nevertheless, in the beginning of the present century few dimension visualization method was in use as a part 2D/3D node-link diagram [19]. Already at this level of abstraction any user may classify the goal and specify further analytical steps for the research, but unfortunately data scaling became an essential issue.

However, currently used technologies for the visualization data are already causing enormous amount of unpleasantness as high space requirements and extremely high tool deployment cost. Though, existing in the present environment faces us to a new limitation based on larger amounts of data to be visualized in contrast to past imagination issue. Modern effective methods are focused on the representation in specified rooms equipped with widescreen monitors or projectors [20]. But an upcoming era of an augmented reality could solve those problems in the nearest future.

## V. Integration with Augmented Reality

It is well known that human brain vision perception capabilities are limited [21]. Furthermore, handling a visualization process on currently used screens requires high time and health costs. This leads to the need of its proper usage in case of the image interpretation. Nevertheless, modern environment is in the process of flooding with countless amount of wearables [22] and various display devices [23].

Augmented reality displays and helmets [24] seem to be an established research and industrial area for the upcoming decade [25]. The use of it in the visualization area might solve a lot of issues from narrow vision angle, navigation, scaling, etc. For example, an angle problem can be solved by offering a way to have a complete 360-degrees view with a helmet. On the other hand, a solution can be obtained with help of specific widescreen rooms, which by definition involves enormous amounts of spending. Focusing on the combination of dynamic projection and interactive filtering visualization methods, augmented reality devices in combination with motion recognition tools might solve a significant scaling problem especially for multidimensional representations which comes to this area from the architectural one. Speaking more precisely, designers (specialized on 3D-visualization) work with a flat projections in order to produce a visual model [26]. However, the only option to present a final image is in moving around it, thus navigation inside the model seems to be another influential issue [27].

From the big data visualization point of view, *scaling* is a significant issue mainly caused by multidimensional systems where exists a need to delve into a branch of information in order to obtain some specific value or knowledge. Unfortunately, it can not be solved from static point of view. Likewise, integration with motion detection wearables [28] would highly increase such visualization system usability. For example, the additional use of MYO armband [29] may be a key to the interaction with visualized data in the most native way. Similar comparison may be given as a *pencil-case* in which one tries to find a sharpener and spreads stationery with his/her fingers.

However, the use of augmented reality displays and helmets is also limited by specific characteristics of the human eye (visual system), such as field of view, scotoma and blind spots [30]. Anyhow, Central vision [31] is the most significant and necessary for human activities such as reading or driving. Additionally, it is responsible for accurate vision in the pointed direction and takes most of the visual cortex in the brain but its retinal size is less than $1\%$ [32]. Furthermore, it captures only two degrees of the vision field which stays the most considerable for text and object recognition. Nevertheless, it is supported with the Peripheral vision which one is responsible for events outside the center of gaze.

Additionally, there exists a research area focusing on the human eye movement patterns during the perception of scenes and objects. However, it can be based on different factors starting from particular culture peculiar properties [33] and up to specific search tasks [34] being in high demand for big data visualization purposes.

Further studies shall be focused on the usage of ophthalmology and neurology for the development of the new visualization tools. Basically, such cross-discipline collaboration would support decision making for the image position selection, which is mainly related to the problem of the significant information losses due to the vision angle extension. Moreover, it is highly important to take in account current hardware quality and screens resolution in addition to the software part. Nevertheless, there is a need of the improvement for multicore GPU processors besides the address bus throughput refinement between CPU and GPU or even replacement for wireless transfer computations on cluster systems.

## VI. Conclusion

Based on the above, it can be concluded that data visualization methodology may be improved by placing the most essential data in the most recognizable area of the human visual field especially for augmented helmets where a head is fixed in a specific position. Moreover, extending it with functions to exclude blind spots and decreased vision sectors would highly improve recognition time for people with such a disease. Furthermore, a step to the wireless solutions would extend the device battery life in addition to the computation and quality improvements.

## References


[1] T. White, *Hadoop: The Definitive Guide, Third Edition*. O'Reilly Media, Inc, 2012.

[2] D. Kuznetzky, "What is big data?," *Virtually Speaking*, Februay 2010.

[3] M. Kaur, "Challenges and issues during visualization of big data," *International Journal For Technological Research In Engineering*, vol. 1, pp. 174–176, 2013.

[4] H. Childs, B. Geveci, W. Schroeder, J. Meredith, K. Moreland, C. Sewell, T. Kuhlen, and E. W. Bethel, "Research challenges for visualization software," *VISUALIZATION GROUP*, 2013.

[5] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," *McKinsey Global Institute*, 2011.

[6] M. Friendly, "Milestones in the history of thematic cartography, statistical graphics, and data visualization," August 2009.

[7] M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Moller, "Visualization task performance with 2d, 3d, and combination displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, January 2006.

[8] S. R. M. Oliveria and O. R. Zaiane, "Geometric data transformation for privacy preserving clustering," *DEPARTMENT OF COMPUTING SCIENCE*, 2003.

[9] C. G. Healey and J. T. Enns, "Large dataasdat at a glance: Combining textures and colors in scientific visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 5, no. 2, 1999.

[10] D. A. Keim, "Designing pixel-oriented visualization techniques: Theory and applications," *IEEE Transactions on visualization and computer graphics*, vol. 6, no. 1, 2000.

[11] M. Kamel and A. Camphilho, "Hierarchic image classification visualization," *ICIAR*, 2013.

[12] A. Buja, D. C. nd Daniel Asimov, and C. Hurley, "Theory of dynemic projection in high-dimensional data visualization," March 2004.

[13] A. Meijester, M. A. Westenberg, and M. H. F. Wilkinson, "Interactive shape preserving filtering and visualization of volumetric data," *Signal and Image Processing*, August 2002.

[14] I. Borg and P. Groenen, "Modern multidimensional scaling: theory and applications (2nd ed.)," *New-York: Springer-Verlag*, 2005.

[15] C. Bajaj and B. Krishnamurthy, "Data visualization techniques," *Trends in Software Ser.*, vol. 6, 1999.

[16] A. Dasgupta, "The role of visualization in the big data era: An end to a means or a means to an end?," *VIS*, 2013.

[17] S. Suthaharan, "Big data classification: Problems and challenges in network intrusion prediction with machine learning," *ACM SIGMETRICS*, 2013.

[18] E. R. Tufte, *The Visual Display for Quantitative Information*. Graphics Press, Chelshire, Connecticut, 1983.

[19] C. von Mering, "Traditional node-link diagram of a network of yeast protein-protein and protein-dna interactions with over 3,000 nodes and 6,800 links.," *Nature*, vol. 417, pp. 399–403, 2002.

[20] J. Aurisano and M. Lewis, "http://grad.uic.edu/," *Electronic Visualization Laboratory at UIC*, 2014.

[21] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: Evidence for a local "association field"," *Vision Res.*, vol. 33, pp. 173–193, 1993.

[22] R. Picard and J. Healy, "Affective wearables," *Springer*, 1997.

[23] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, "Augmented reality technologies, systems and applications," *Springer*, vol. 51, pp. 341–377, 2010.

[24] B. Chacos, "Shining light on virtual reality: Busting the 5 most inaccurate oculus rift myths," *PCWorld*, March 2014.

[25] D. van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *The International Journal of Virtual Reality*, vol. 9, pp. 1–20, 2010.

[26] J. L. Pina, E. Cerezo, and F. Seron, "Semantic visualization of 3d urban environments," *Multimed Tools Applications*, vol. 59, pp. 505–521, 2012.

[27] D. Fonseca, S. Villagrasa, N. Marta, E. Redondo, and A. Sanchez, "Visualization methods in architecture education using 3d virtual models and augmented reality in mobile and social networks," *Procedia - Social and Behavioral Sciences*, vol. 93, pp. 1337–1343, 2013.

[28] J. P. Varkey, D. Pompili, and T. A. Walls, "Human motion recognition using a wireless sensor-based wearable system," *Springer*, 2011.

[29] MYO, "https://www.thalmic.com/en/myo/."

[30] P. J. Foster, R. Buhrmann, H. A. Quigley, and G. J. Johnson, "The definition and classification of glaucoma in prevalence surveys," *Br J Ophthalmol*, vol. 86, pp. 238–242, 2002.

[31] M. F. Deering, "The limits of human vision," *2nd International Immersive Projection Technology Workshop*, 1998.

[32] J. H. Krantz, "Chapter 3: The stimulus and anatomy of the visual system," *Experiencing Sensation and Perception. Pearson Education. ISBN 978-0-13-097793-9. OCLC 711948862.*, 6 April 2012.

[33] H. F. Chua, J. E. Boland, and R. E. Nisbett, "Cultural variation in eye movements during scene perception," *PNA*, vol. 102 no. 35, pp. 12629–12633, 2005.

[34] G. J. Zelinsky, H. Adeli, Y. Peng, and D. Samaras, "Modelling eye movements in a categorical search task," *Phisiolophical Transactions of The Royal Society*, September 2013.