فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

پیاده سازی نشان دادن اعداد بوسیله انگشتان دو دست بصورت همزمان گفتار و نوشتار فارسی

امیرحسین دنیائی گزکوه[1]، الیاس عرب محقی[2]

[1]دانشجوی کارشناسی مهندسی کامپیوتر نرم افزار ، دانشگاه ملی و مهارت شهید محمد منتظری مشهد
cactuspydon@gmail.com
[2]گروه مهندسی کامپیوتر، دانشگاه ملی و مهارت شهید محمد منتظری مشهد
Elyas.arabmoh@gmail.com

چکیده

این مقاله به معرفی یک سامانه یا سیستم هوشمند چندمنظوره است که به‌طور همزمان از دو فناوری پیشرفته در حوزه‌های پردازش تصویر و پردازش صوت بهره می‌برد. هدف این پروژه، توسعه‌ی ابزاری است که قادر به شناسایی صدا و شمارش انگشتان دست به صورت همزمان باشد. این سیستم از یک طرف با استفاده از دوربین وب‌کم، قادر به شناسایی و شمارش انگشتان دست کاربر است و از طرف دیگر، با استفاده از میکروفن، گفتار فارسی کاربر را به متن تبدیل می‌کند. بخش پردازش تصویر با استفاده از کتابخانه آن برای شناسایی و پیگیری حرکات دست ها و انگشتان طراحی شده است و بخش کتابخانه ای پردازش صوت، گفتار فارسی کاربر را به متن تبدیل می کند که اهداف آن شناسایی صدا به زبان فارسی،شناسایی اعداد نشان داده شده انگشتان دست،و ترکیب همزمان این دو نوع شناسایی می باشند. در روش تحقیق از کتابخانه ها و ابزارهای نرم افزاری و سخت افزاری مناسب برای بدست آمدن نتیجه استفاده شده است که دست آوردهای این تحقیق در زمینه هایی مانند دستورات صوتی و تعاملات بدون لمس کاربرد داشته باشد.

واژه‌های کلیدی: شناسایی صدا،شناسایی صوت،پردازش تصویر، گفتار فارسی،تعاملات بدون لمس

د.وره: چهارم
شماره: چهار
آدرس سایت:nrbste.ir
ایمیل:nrbste@gmail.com

دریافت مقاله:1403/10/01
اصلاحیه مقاله:-
پذیرش مقاله:1403/10/11
انتشار مقاله:1403/10/13

فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

مقدمه

معرفی پروژه

پروژه‌ی حاضر یک سیستم هوشمند چندمنظوره است که به‌طور همزمان از دو فناوری پیشرفته در حوزه‌های پردازش تصویر و پردازش صوت بهره می‌برد. هدف این پروژه، توسعه‌ی ابزاری است که قادر به شناسایی صدا و شمارش انگشتان دست به صورت همزمان باشد. این سیستم از یک طرف با استفاده از دوربین وب‌کم، قادر به شناسایی و شمارش انگشتان دست کاربر است و از طرف دیگر، با استفاده از میکروفن، گفتار فارسی کاربر را به متن تبدیل می‌کند.

در بخش پردازش تصویر، از کتابخانه‌ی MediaPipe برای شناسایی نقاط کلیدی دست و انگشتان آن استفاده شده است. این کتابخانه به‌طور خاص برای شناسایی و پیگیری حرکات دست‌ها و انگشتان طراحی شده است و با الگوریتم‌های پیشرفته خود، توانایی شناسایی تعداد انگشتان دست کاربر را در لحظه به‌دست می‌آورد. این اطلاعات به‌طور زنده بر روی صفحه نمایش داده می‌شود، که می‌تواند برای کاربردهایی مانند سیستم‌های آموزشی، بازی‌های تعاملی، یا حتی ابزارهای تشخیص حرکت دست در محیط‌های واقعیت افزوده استفاده شود.

در بخش پردازش صوت، از کتابخانه‌ی SpeechRecognition استفاده شده است تا گفتار فارسی کاربر را به متن تبدیل کند. این فناوری قادر است صدای کاربر را از طریق میکروفن دریافت کرده و آن را به‌طور دقیق به متنی قابل استفاده برای سیستم تبدیل نماید. این ویژگی می‌تواند در کاربردهای مختلفی نظیر تعامل صوتی با سیستم‌ها، تبدیل گفتار به متن برای افرادی که نمی‌توانند تایپ کنند، یا به عنوان یک سیستم کنترل صوتی برای دستگاه‌های مختلف به کار گرفته شود.

ترکیب این دو فناوری (پردازش تصویر و صوت) یک سیستم هوشمند و تعاملی را ایجاد می‌کند که توانایی برقراری ارتباط با کاربر را از طریق دو کانال مختلف (صوتی و بصری) دارد. این پروژه می‌تواند در بسیاری از حوزه‌ها مفید باشد:

ابزارهای تعاملی :سیستم‌هایی که به کاربر امکان می‌دهند تا با استفاده از صدا و حرکت دست، تعامل با دستگاه‌ها و برنامه‌ها را انجام دهند.

آموزش : برنامه‌های آموزشی که نیاز به تعامل بصری و صوتی با کاربران دارند، مانند آموزش زبان اشاره یا ابزارهای آموزشی برای کودکان.

سیستم‌های کمکی: به‌ویژه برای افراد دارای نیازهای خاص، مانند کسانی که مشکلات حرکتی یا شنوایی دارند، این برنامه می‌تواند به‌عنوان یک ابزار کمکی برای تعامل آسان‌تر با محیط‌های دیجیتال استفاده شود.

این پروژه نمونه‌ای از کاربرد هوش مصنوعی است که از ترکیب دو حوزه‌ی مختلف پردازش صوت و پردازش تصویر برای ایجاد یک سیستم هوشمند و چندمنظوره استفاده می‌کند. این پروژه همچنین نشان می‌دهد که چگونه می‌توان از تکنولوژی‌های مدرن برای حل مشکلات واقعی و بهبود دسترسی و تعامل کاربران در دنیای دیجیتال بهره‌برداری کرد.

اهداف پروژه

۱. شناسایی صدا به زبان فارسی: هدف اصلی این پروژه توسعه یک سیستم شناسایی صدا است که قادر به دریافت و پردازش گفتار به زبان فارسی باشد. این سیستم باید قادر به شناسایی و تبدیل گفتار به متن به‌طور دقیق و سریع باشد.

۲. شمارش انگشتان دست: توسعه الگوریتمی که توانایی شناسایی دست‌ها و شمارش انگشتان آن‌ها را از طریق دوربین وب‌کم فراهم کند. این قابلیت می‌تواند برای کاربردهای مختلفی مانند تعامل‌های مبتنی بر دست یا کنترل‌های بدون لمس مفید باشد.
۳. ترکیب همزمان شناسایی صدا و پردازش تصویر: هدف دیگر این پروژه ایجاد یک سیستم است که قادر به انجام هر دو عمل شناسایی صدا و شمارش انگشتان به‌طور همزمان باشد، بدون اینکه عملکرد هر کدام بر دیگری تاثیر منفی بگذارد.
۴. ایجاد تجربه کاربری تعاملی و کاربردی: پروژه به‌دنبال ایجاد یک ابزار مفید است که می‌تواند به‌صورت تعاملی با کاربر در ارتباط باشد و از این طریق تجربه کاربری بهتری را فراهم کند.

روش تحقیق و ضرورت‌های پروژه

افزایش تعاملات بدون تماس، در دنیای امروز که شیوع بیماری‌ها و نگرانی‌های بهداشتی افزایش یافته است، استفاده از سیستم‌هایی که نیاز به تماس فیزیکی کمتری دارند، ضروری است. این پروژه می‌تواند به کاربران این امکان را بدهد که از طریق حرکت دست یا فرمان‌های صوتی بدون لمس دستگاه‌ها با سیستم‌ها تعامل داشته باشند. کاربرد در محیط‌های مختلف ،این فناوری می‌تواند در صنایع مختلف از جمله آموزش، پزشکی، امنیت و حتی سرگرمی کاربرد داشته باشد. به‌ویژه در محیط‌های آموزشی، دانش‌آموزان می‌توانند از آن برای کنترل محتوا و تعامل با سیستم‌های آموزشی استفاده کنند. پشتیبانی از زبان فارسی، یکی از ویژگی‌های منحصر به‌فرد این پروژه، پشتیبانی از زبان فارسی است که در بسیاری از پروژه‌های مشابه کمتر دیده می‌شود. این ویژگی می‌تواند برای کاربران فارسی‌زبان بسیار مفید باشد.توسعه فناوری‌های هوش مصنوعی، پروژه به‌طور غیرمستقیم به پیشرفت در زمینه‌های پردازش زبان طبیعی (NLP) و بینایی کامپیوتری (Computer Vision) کمک می‌کند و می‌تواند زمینه‌ساز توسعه‌های بیشتر در این حوزه‌ها باشد.

فناوری‌ها و ابزارهای استفاده‌شده

این پروژه از ترکیبی از فناوری‌های پیشرفته در زمینه پردازش صدا، بینایی کامپیوتری، و برنامه‌نویسی استفاده کرده است. ابزارها و کتابخانه‌های زیر برای دستیابی به اهداف پروژه انتخاب شده‌اند:

کتابخانه‌ها و ابزارهای پردازش صدا

SpeechRecognition: این کتابخانه برای شناسایی صدا و تبدیل آن به متن استفاده می‌شود. قابلیت پشتیبانی از زبان فارسی و دقت بالا در شناسایی گفتار از ویژگی‌های برجسته این ابزار است.

Microphone: برای دریافت ورودی صدا از میکروفن دستگاه استفاده می‌شود که بخشی از ابزارهای مرتبط با `speech_recognition` است.

کتابخانه‌ها و ابزارهای پردازش تصویر

Mediapipe: کتابخانه‌ای قدرتمند و متن‌باز برای پردازش حرکات بدن، دست‌ها، چهره، و دیگر بخش‌های انسانی است. در این پروژه برای شناسایی دست‌ها و انگشتان و انجام پردازش‌های موردنیاز استفاده شده است.

OpenCV: این کتابخانه برای پردازش و نمایش ویدئو به کار گرفته شده است. از OpenCV برای دریافت فریم‌های ویدئویی از وب‌کم، تبدیل تصاویر به فرمت‌های مختلف، و نمایش نتایج استفاده می‌شود.

فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

زبان و ابزارهای برنامه‌نویسی

Python: زبان برنامه‌نویسی اصلی این پروژه به دلیل انعطاف‌پذیری، سادگی، و پشتیبانی قوی از کتابخانه‌های هوش مصنوعی و پردازش تصویر انتخاب شده است.

Threading: ماژول `threading` برای اجرای همزمان شناسایی صدا و شمارش انگشتان به کار رفته است. این قابلیت امکان انجام چند وظیفه‌ای را بدون تأخیر یا اختلال فراهم می‌کند.

ابزارهای سخت‌افزاری

وب‌کم: برای دریافت تصاویر و ویدئوهای زنده از دست‌ها استفاده شده است.

میکروفن: برای دریافت صدای کاربر و تبدیل آن به متن به کار می‌رود.

ویژگی‌ها و قابلیت‌های کلیدی

پردازش همزمان صدا و تصویر: این پروژه از threading برای اجرای دو وظیفه به طور همزمان بهره می‌برد.

پشتیبانی از زبان فارسی: شناسایی صدا به زبان فارسی یکی از ویژگی‌های مهم پروژه است که آن را برای کاربران فارسی‌زبان کاربردی‌تر می‌کند.

واکنش در لحظه: سیستم به صورت بلادرنگ نتایج را پردازش و به کاربر نمایش می‌دهد. این ترکیب از فناوری‌ها و ابزارها، پروژه را به یک ابزار چندمنظوره و کاربردی تبدیل کرده است.

کتابخانه‌ها و ابزارهای مورد استفاده

SpeechRecognition: برای شناسایی صدا و تبدیل آن به متن.

PyAudio: برای دریافت صدا از میکروفن.

Mediapipe: برای شناسایی دست‌ها و شمارش انگشتان.

OpenCV: برای پردازش تصاویر و دریافت ویدئو از وب‌کم.

Threading: برای اجرای همزمان وظایف (پیش‌فرض در پایتون)

Numpy: برای پردازش داده‌های عددی

```
pip install SpeechRecognition pyaudio mediapipe opencv-python numpy
```

روش کارکرد برنامه

این برنامه دو عملیات اصلی را به طور همزمان انجام می‌دهد: شناسایی صدا و شمارش انگشتان دست. در ادامه، مراحل اجرای هر بخش توضیح داده شده است:

```
pip install SpeechRecognition pyaudio mediapipe opencv-python numpy
```

فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

شناسایی صدا (Speech Recognition)

الف) راه‌اندازی میکروفن: ابتدا میکروفن برای دریافت صدا از محیط آماده می‌شود.

ب) پردازش صدا: برنامه به‌طور مداوم به صدا گوش می‌دهد و وقتی صدا دریافت می‌شود، آن را به متن فارسی تبدیل می‌کند.

ج) نمایش متن: متن شناسایی‌شده به کاربر نمایش داده می‌شود.

د) مدیریت خطا: اگر صدا قابل شناسایی نباشد، پیام خطا نمایش داده می‌شود.

شمارش انگشتان دست (Hand Finger Counting)

دریافت ویدئو از وب‌کم: ویدئو از وب‌کم دریافت شده و به‌طور مداوم پردازش می‌شود.

شناسایی دست‌ها: از کتابخانه Mediapipe برای شناسایی دست‌ها و نقاط مختلف انگشتان استفاده می‌شود.

شمارش انگشتان: با استفاده از نقاط خاص دست، انگشتان شناسایی شده و تعداد آن‌ها شمارش می‌شود.

نمایش نتایج: تعداد انگشتان هر دست و مجموع آن‌ها روی ویدئو نمایش داده می‌شود.

اجرای همزمان دو عملیات

برای انجام این دو وظیفه به‌طور همزمان، از ماژول `threading` پایتون استفاده شده است. این امکان به برنامه می‌دهد که همزمان به صدا گوش دهد و شمارش انگشتان را انجام دهد، بدون اینکه یکی از آن‌ها بر دیگری تأثیر بگذارد.این برنامه در نهایت به‌طور بلادرنگ به تعاملات صوتی و تصویری پاسخ می‌دهد.

کد بخش شناسایی صدا:

```
def recognize_speech():
    recognizer = sr.Recognizer()
    mic = sr.Microphone()
    print("\n🎤 Speech Recognition Initialized: Say something in Persian!\n")
```

این بخش در داخل تابع recognize_speech() قرار دارد که با استفاده از کتابخانه SpeechRecognition و میکروفن، به‌طور مداوم صدا را ضبط کرده و آن را به متن تبدیل می‌کند.

ضبط صدا:

در اینجا، میکروفن برای ضبط صدا به‌طور پیوسته فعال می‌شود. تنظیمات مربوط به نویز محیط نیز در اینجا انجام می‌شود

```
with mic as source:
    recognizer.adjust_for_ambient_noise(source)
    print("🎙️ Listening... Please speak now.")
    audio = recognizer.listen(source)
```

شناسایی متن از صدا:

```
text = recognizer.recognize_google(audio, language='fa-IR')
print(f"\n✅ Recognized Text: {text}\n")
```

فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

در این قسمت، از سرویس Google Speech Recognition برای شناسایی صدا به زبان فارسی استفاده می‌شود و متن استخراج‌شده به کاربر نمایش داده می‌شود.

راه‌اندازی وب‌کم:

```
cap = cv2.VideoCapture(0)
# Set width to 1280 for larger display
cap.set(cv2.CAP_PROP_FRAME_WIDTH, 1280)
# Set height to 720 for larger display
cap.set(cv2.CAP_PROP_FRAME_HEIGHT, 720)

print("\n✋ Finger Counting Initialized: Press 'Q' to exit.\n")
```

در اینجا، وب‌کم برای دریافت ویدئو از محیط راه‌اندازی می‌شود و اندازه فریم ویدئو تنظیم می‌شود.

پردازش تصویر و شناسایی دست‌ها:

```
# Convert the frame to RGB
frame_rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
result = hands.process(frame_rgb)
```

در اینجا، فریم ویدئو به فرمت RGB تبدیل می‌شود و سپس پردازش برای شناسایی دست‌ها انجام می‌شود. از mediapipe برای شناسایی نقاط مختلف دست‌ها استفاده می‌شود.

شمارش انگشتان دست‌ها:

```
# Check if thumb is open (special case due to its orientation)
if hand_landmarks.landmark[4].x < hand_landmarks.landmark[3].x if hand_landmarks.landmark[4].y < hand_landmarks.landmark[3].y else hand_landmarks.landmark[4].x > hand_landmarks.landmark[3].x:
    finger_count += 1

# Check other fingers
for tip, joint in zip(finger_tips[1:], lower_joints[1:]):
    if hand_landmarks.landmark[tip].y < hand_landmarks.landmark[joint].y:
        finger_count += 1
```

در اینجا، انگشتان دست‌های شناسایی‌شده شمارش می‌شوند. برای هر دست، نقاط انتهایی انگشت‌ها (انگشت شست و انگشتان دیگر) بررسی می‌شود تا تعداد انگشتان بازشده شمارش گردد.

نمایش نتایج شمارش انگشتان:

```
cv2.putText(frame, f"Total Fingers: {finger_count_total}", (
    20, 50), cv2.FONT_HERSHEY_SIMPLEX, 1.5, text_color, 3)
```

این کد برای نمایش تعداد کل انگشتان شمارش‌شده در بالای تصویر استفاده می‌شود.

فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

ایجاد و اجرای (Threads)

با

```
# Running both tasks simultaneously using threading
speech_thread = threading.Thread(target=recognize_speech, daemon=True)
finger_thread = threading.Thread(target=count_fingers, daemon=True)

speech_thread.start()
finger_thread.start()

speech_thread.join()
finger_thread.join()
```

استفاده از کتابخانه threading، این بخش به‌طور همزمان هر دو عملیات شناسایی صدا و شمارش انگشتان را اجرا می‌کند. در اینجا، دو (Thread) به‌طور همزمان ایجاد و شروع به اجرا می‌کنند: یکی برای شناسایی صدا و دیگری برای شمارش انگشتان دست. از ویژگی daemon=True برای این استفاده می‌شود که نخ‌ها به‌طور خودکار وقتی که برنامه اصلی تمام شود، متوقف شوند.

نمایش متن و اطلاعات:

```
# Show exit instructions
cv2.putText(frame, "Press 'Q' to exit", (20,
            frame.shape[0] - 20), cv2.FONT_HERSHEY_SIMPLEX, 0.8, (0, 255, 0), 2)
```

این کد

برای نمایش دستور خروج از برنامه استفاده می‌شود.

نمایش تصاویر و فریم‌های ویدئویی:

در

```
cv2.imshow("Finger Count", frame)
```

اینجا، تصویر ویدئو با شمارش انگشتان روی صفحه نمایش داده می‌شود

راهنمای استفاده از برنامه

نصب وابستگی‌ها: ابتدا کتابخانه‌های لازم را نصب کنید.

اجرای برنامه: برای اجرای برنامه، کد را در محیط Python اجرا کنید.

عملکرد شناسایی صدا:

میکروفن باید به دستگاه متصل باشد و به‌درستی تنظیم شده باشد.

پس از اجرای برنامه، میکروفن به‌طور مداوم به صدا گوش می‌دهد. برنامه پس از شناسایی صدای شما، متن آن را به نمایش در می‌آورد.

شما می‌توانید به‌صورت فارسی صحبت کنید، و برنامه به‌طور خودکار متن را شناسایی کرده و در کنسول نمایش می‌دهد.

اگر صدا به‌طور واضح و بدون نویز اضافی ضبط نشود، ممکن است برنامه نتواند صدای شما را شناسایی کند.

عملکرد شمارش انگشتان دست:

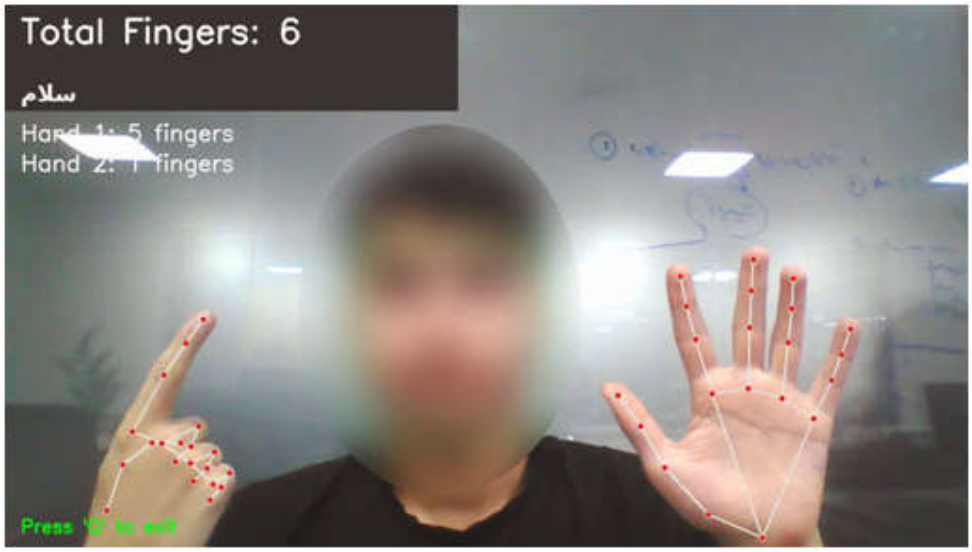
دوربین وب‌کم باید فعال و در دسترس باشد.

برنامه به‌طور مداوم تصویر دریافتی از دوربین را پردازش می‌کند و تعداد انگشتان هر دست را به‌طور زنده روی صفحه نمایش می‌دهد.
برنامه از کتابخانه MediaPipeبرای شناسایی دست و شمارش انگشتان استفاده می‌کند. این قابلیت به‌طور طبیعی در محیط‌های روشن‌تر بهتر عمل می‌کند.
هنگامی که دست شما در میدان دید دوربین قرار گیرد، سیستم انگشتان شما را شناسایی کرده و تعداد آن‌ها را به نمایش می‌گذارد.

خروج از برنامه:

برای خروج از برنامه، می‌توانید به‌سادگی کلید 'Q'را فشار دهید. پس از فشردن کلید ,'Q' برنامه به‌طور خودکار از دوربین وب‌کم و میکروفن قطع شده و پنجره‌های OpenCV بسته می‌شوند

اجرای برنامه و شروع کار



شکل ۱) نشان دادن عدد 6

نکات مهم

اطمینان حاصل کنید که میکروفن و دوربین به درستی کار می‌کنند. در غیر این صورت، ممکن است برخی از ویژگی‌ها مانند شناسایی صدا یا شمارش انگشتان دچار اختلال شود.
برنامه به‌طور همزمان از هر دو فناوری شناسایی صدا و پردازش تصویر استفاده می‌کند، بنابراین ممکن است در سیستم‌های با منابع سخت‌افزاری محدود، عملکرد بهینه نباشد.
اگر از سیستم‌عامل‌های مختلف مانندWindows ، macOS یا Linux استفاده می‌کنید، ممکن است نیاز به تنظیمات خاصی برای پیکربندی میکروفن یا دوربین باشد.

این راهنمای استفاده به شما کمک می‌کند تا از امکانات این برنامه به‌طور مؤثر و صحیح استفاده کنید.

نتیجه‌گیری

برنامه ارائه‌شده ترکیبی از دو فناوری پیشرفته است، شناسایی صدا و شمارش انگشتان دست. این برنامه با استفاده از کتابخانه‌های SpeechRecognition و OpenCV همراه با MediaPipe توانسته است عملکردی مفید و کاربردی را ارائه دهد که می‌تواند در بسیاری از زمینه‌ها مانند دستورات صوتی و تعاملات بدون لمس کاربرد داشته باشد. در شناسایی صدا، این بخش از برنامه قادر به تشخیص کلمات و جملات به زبان فارسی است و به‌طور پیوسته به صدای کاربر گوش می‌دهد. و شمارش انگشتان دست، این قسمت از برنامه به‌صورت زنده انگشتان دست کاربر را شناسایی کرده و تعداد انگشتان را به‌دقت نمایش می‌دهد.

با استفاده از threading، این دو فرآیند به‌طور همزمان اجرا می‌شوند، که به کاربر این امکان را می‌دهد تا به‌طور همزمان از هر دو ویژگی استفاده کند. در نهایت، این برنامه می‌تواند در کاربردهایی مانند کنترل صوتی و تعامل بدون لمس، یا در محیط‌های خاص مثل آموزش یا سرگرمی‌های دیجیتال، مفید باشد.

این پروژه نشان می‌دهد که ترکیب فناوری‌های مختلف می‌تواند راه‌حل‌های خلاقانه و جالبی برای مشکلات مختلف فراهم کند.

منابع

[1] Hinton, G., Deng, L., Yu, D., et al., "Deep Neural Networks for Acoustic Modeling in Speech
[2] Bradski, G., Recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2019.
[3] Zhang, Z., "MediaPipe: "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2021.
A Framework for Building Perception Pipelines," *Proceedings of IEEE Conference on Computer*
[4] Raj, S., & Kumar, M., *Vision and Pattern Recognition (CVPR)*, pp. 1234–1240, 2022
"Threading and Concurrent Programming in Python," *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 14, no. 7, pp. 95–101, 2022.

فصلنامه علمی تخصصی پژوهش های نوین بین رشته ای علوم پایه و فنی و مهندسی
Specialized Scientific Quarterly of New Interdisciplinary Researches in Basic Science and Technical and Engineering
« E-ISSN:2980-9061 »

Implementation of showing numbers with the fingers of two hands at the same time in Farsi speech and writing

Amirhosein donyaie gazkooh, Elyas arabmoheghi
Bachelor student of computer software engineering, Shahid Mohammad Montazeri National University of Mashhad
cactuspydon@gmail.com
Department of Computer Engineering, Shahid Mohammad Montazeri National University of Mashhad
Elyas.arabmoh@gmail.com

Abstract- This article introduces a multipurpose intelligent system that simultaneously uses two advanced technologies in the fields of image processing and sound processing. The goal of this project is to develop a tool that can recognize voice and count fingers simultaneously. On the one hand, this system is able to identify and count the user's fingers using a webcam, and on the other hand, using a microphone, it converts the user's Farsi speech into text. The image processing section is designed using its library to identify and track the movements of hands and fingers, and the audio processing library section converts the user's Farsi speech into text, the goals of which are voice recognition in Persian language, recognition of displayed numbers. Fingers and simultaneous combination of these two types of identification. In the research method, appropriate software and hardware libraries and tools have been used to obtain results that the achievements of this research can be used in areas such as voice commands and touch-free interactions.

Keywords: Voice recognition, voice recognition, image processing, Persian (Farsi) speech, non-touch interactions