

Artificial Intelligence enabled virtual sixth sense application for the disabled

Keerthana V , Sangeetha P, Aishwarya N, Aishwarya B

Abstract

The sixth sense is a multi-platform app for aiding the people in need that is people who are handicapped in the form of lack of speech (dumb), lack of hearing (deaf), lack of sight (blind), lack of judicial power to differentiate between objects (visual agnosia) and people suffering from autism (characterized by great difficulty in communicating and forming relationships with other people and in using language and abstract concepts). Our current implementation of the product is on two platforms, namely, mobile and a web app. The mobile app even works for object detection cases in offline mode. What we want to achieve using this is to make a better world for the people suffering from disabilities as well as an educational end for people with cognitive disabilities using our app. The current implementation deals with object recognition and text to speech and a speech to text converter. The speech to text converter and text to speech converter utilized the Web Speech API (Application Program Interface) for the website and text to speech and speech to text library for the mobile platform. The object recognition wouldn't fetch enough use out of a website. Hence, it has been implemented on the mobile app utilizing the Firebase ML toolkit and different pre-trained models, which are both available offline as well as online.

Keywords: Sixth sense, disabilities, Web Speech API, Firebase ML toolkit, cognitive disabilities

1. Introduction

In the past couple of decades, technology has been at the forefront of every revolutionary change that is occurring at both global and local levels. These techniques have led to make the lives of humans much more simplified, smooth, and even comfortable at different levels. From the age of the internet to the age of electric vehicles, technology and especially engineering has led to inventions and major level improvements on each spectrum imaginable and even beyond that. Especially in the past decade, the widespread use of the internet along with personal computing devices has also brought major changes to the lifestyle of the people around the globe. From getting daily news updates to booking flights to even groceries to our doorstep, the world has come a long way through the use of technology. The new spectrum under research in technology is artificial intelligence, mostly abbreviated to AI. The world does not want just machines to do what they are told but even expect devices to work like us. Machine Learning, a subsection of AI, is the hottest technology right now being implemented on a daily basis and is supposed to reach its peak in the next decade. Now, as the world has become a better place by providing everything at ease through technology to humans, we need to utilize the technology for differently-abled people. These people are blind, deaf, dumb, or those who suffer from cognitive disabilities. The main aim is to make the internet easily accessible to differently-abled people, make the internet as easy to use as to others, as comfortable as it is to others. To make their daily lives simpler through technology. The technology to be used here would be object recognition, Optical Character Recognition, text to speech

speech to text conversions. These could be used in different forms. Object Recognition] is great computer technology. Humans can easily detect and identify objects present in an image. The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects and detect obstacles with little conscious thought.

These both can also be applied even to classroom learning, providing students with interactive learning for students[13] who are blind to have a sense of their books and even what is being taught. For people who are blind or dumb are not able to easily use their devices with ease. They could utilize their devices like computers and mobile phones in a much better way if they could interact with these devices in the way they are at ease. Blind people [1] would be much at ease to utilize their devices by providing speech commands to devices rather than typing and scrolling and dumb people would be able to use text format for interaction with others through the help of their device rather than sign language, which are time-consuming and even not known to everyone [11]. The converter of speech to text and text to speech[12] would make the interaction with the devices and others for blind and dumb people better[18] with a conclusion.

2. Motivation

Braille Script was a remarkable invention in the 17th century for the blind; it provided the people with disabilities to read without the conventional method of eyes rather than through the sense of touch. So this was the perfect example, as things came per se the books, they were then tweaked for people with disabilities so they can also benefit from reading at their comfort and through their senses. The last couple of decades have been remarkable in the revolution through computer science and its technology. The world has become more tightly bound to each other, even with the distances among them, comfort has increased, and the lifestyle has become more technology-driven than ever before. These benefits should also reach the differently-abled people. They should also be able to have a life at their comfort through the use of technology. They should also be able to use their devices at ease. This is the major motivation for this project. An event that brought this thinking was when once we saw a blind person was unable to cross the road, and no one was helping him, we went ahead helped him; then he even said "it would have been better than in this modern age something for use would be made so that I won't require any assistance," then a thought came, that Google maps have led to provide people with directions without any assistance[3], it has been the best thing people on the road could

have. Now, why is there not any such technology developed so that blind people[1] can walk around with a sense of their surroundings, what is the way to move, is it safe to walk right now, if there is any obstruction ahead or not. Hence, the motivation for this project is to make technology reach every person, and at their comfort, technology shall change as per the person, not the other way round. Therefore, This needs to be achieved, and this is what is the driving force behind this project.

3. Different application platforms

Using Android smartphones and web-based solutions [14-18], we can implement speech to text and text to speech systems. Using Android smartphones, we approached the solution to solve

the three constituents. For object recognition, we used the Firebase ML Kit for the purpose of object tracking, recognition, and classification. When a captured image is sent to an ML kit, it returns back a list of maximum five detected objects. Each of these is provided an ID, and an object list is returned. This can also work in video format apart from images. An option for coarse classification is also available. Now from the list, the first object in the list is displayed, which is the prominent object. For Text Recognition, the ML OCR kit part was used. If this image is sent to an online or offline mode, it detects the image and all text parts to be analyzed on the basis of its font and size. A list is then returned of objects, and the prominent role is displayed.

Web

Using a web page, we approached the solution of solving the three constituents of the problem statement: For Text-To-Speech, we used Google Web toolkit TTS Which synthesizes speech from the text for immediate playback or to create a sound file. A TextToSpeech instance can only be used to synthesize text once it has completed its initialization. We can define the rate of speech, volume, and even pitch of the voice synthesizer. Currently added the support for around 19 languages.

Result

Text to Speech Converter was implemented on both mobile and web apps with an accuracy upwards of 90% and along with the support of multiple languages. Speech to Text Converter was implemented on both mobile and web apps with an accuracy of around 85% and along with the support of multiple languages input. Object detection was also implemented using the firebase ML kit with an accuracy of 80%. The threshold has been set to 0.7f, which makes the result at least 70% accurate. Firebase provides the backend as a service. Object Character Recognition also was implemented with a similar accuracy of 85%. Both of these were performed only with the image as input for this detection and recognition.

Conclusion and Future Scope

Implemented the object tracking, recognition & classification, and character recognition in offline mode and guarded the app to shrink the size of the app. The main highlight of the project hence came forward, which was being an application that provided a one-stop-shop solution to all the sections of differently-abled people. Integration has provided a seamless User interface/experience for the initial setup. Another point achieved here was no extra hardware; hence, no additional cost to utilize the service. The application still does depend on the camera picture quality for object detection and OCR but is still high enough in confidence level (70%) for most of the cases which were covered. Higher accuracy could be achieved in the future scope of the implementation through the use of custom models for object detection and text recognition as it could take into account the cases of objects for differently-abled people and work on those only yielding faster and accurate results. Funding: "This research received no external funding" Conflicts of Interest: "The authors declare no conflict of interest."

References

- [1] Bigham, J. P., Jayant, C., Miller, A., White, B., & Yeh, T. (2010, June). VizWiz:: Locatelt-enabling blind people to locate objects in their environment. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops (pp. 65-72). IEEE. Vision and Pattern Recognition-Workshops (pp. 65-72). IEEE.
- [2] Manduchi, R., Kurniawan, S., & Bagherinia, H. (2010, October). Blind guidance using mobile computer vision: A usability study. In Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (pp. 241-242).
- [3] Ivanchenko, V., Coughlan, J., Gerrey, W., & Shen, H. (2008, October). Computer vision-based clear path guidance for blind wheelchair users. In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (pp. 291-292).
- [4] Johnsen, A., Grønli, T. M., & Bygstad, B. (2012). Making touch-based mobile phones accessible for the visually impaired. Norsk informatikkonferanse,(Bodø, Norway, 2012).
- [5] Matusiak, K., Skulimowski, P., & Strumiłło, P. (2013, June). Object recognition in a mobile phone application for visually impaired users. In 2013 6th International Conference on Human System Interactions (HSI) (pp. 479-484). IEEE.
- [6] Hermus, K., & Wambacq, P. (2006). A review of signal subspace speech enhancement and its application to noise robust speech recognition. EURASIP Journal on Advances in Signal Processing, 2007(1), 045821.
- [7] Dimitrov, V., Jullien, G., & Muscedere, R. (2017). Multiple-base number system: theory and applications. CRC press.
- [8] Huyan, Z., Xu, L., Fang, S., Liu, Z., Zhang, X., & Li, L. (2014). Field information acquisition system research based on offline speech recognition. Int. J. Database Theory Appl, 7, 45-58.
- [9] Omankhanlen, A. E., & Ogaga-Oghene, J. (2013). The Dynamics of Global Strategy and Strategic Alliances in International Trade and Investment. INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT, 3(12), 41-48.
- [10] Kamble, K., & Kagalkar, R. (2014). A review: translation of text to speech conversion for Hindi language. International Journal of Science and Research (IJSR) Volume, 3.
- [11] Arora, S. J., & Singh, R. P. (2012). Automatic speech recognition: a review. International Journal of Computer Applications, 60(9).
- [12] BELGHIT, H., & BELLARABI, A. Object Recognition Based on ORB Descriptor for Markerless Augmented Reality.
- [13] Gill, J. (2000). Personal electronic mobility devices. Information for Professionals Working with Visually Disabled People. <http://www.tiresias.org>.
- [14] Chen, C., & Raman, T. V. (2009). Announcing eyes-free shell for Android. Retrieved December 21, 2016.
- [15] Coughlan, J., & Manduchi, R. (2009). Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. International Journal on Artificial Intelligence Tools, 18(03), 379-397.

