

Virtual Assistance for Visually Impaired

Abhijna UJ¹, Dharani K², Kavya K³, Roja G⁴, Prof. Kavitha Juliet⁵

Students, Department of Computer Science^{1,2,3,4}

Assistant Professor, Department of Computer Science⁵

Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari, Karnataka, India

abhijnauj.cse.rymec@gmail.com, dharanidharu55@gmail.com, kavya.cse.rymec@gmail.com,
rojan721.in@gmail.com.

Abstract: *The purpose of the project is to develop a system/assistant that will assist a visually impaired person and communicate the person by speaking through the earpiece. As it has been noted from time to time that the way world is developing faster in terms of technology by creating the robot or AI's in replacement of human resource as much as it has it's negative impact it also has it's positive impact. In case of artificial intelligence where virtual assistance is created for aiding the people and mainly it helpful for special aided people in their day to day work such as recognizing a person or distinguishing an object, these tasks are straightforward for common people but can be very difficult for people that are visually impaired or visionless. The system will help the person to recognize people, add new faces and detect objects that are in their vicinity and their lives can be made smoother by assisting them to descry what is present in front of them at that instant. The aim of the project is to develop a mobile application that consist of numerous deep learning models that will help applications increase their administration. The primary working of the system will consist of the camera where it'll be continuously feeding the images for inputs, the core system processing this input information and the earpiece will act as the output device to provide this result to the user. This project applies the concept of Deep learning i.e. Neural networks. The system comprises a camera that acquires images and sends them to the application, where a powerful processor derives information from them and explains them to the user through a distinct audible message.*

Keywords: Deep learning, Neural networks, core system

I. INTRODUCTION

“Virtual Assistant for visually impaired”, this project applies the concept of Deep learning i.e. Neural networks. The models employed for our project are - Face Detection and Object Detection. The system comprises a camera that acquires images and sends them to the application, where a powerful processor derives information from them and explains them to the user through a distinct audible message. The device will continuously detect all the faces in front of the person and verify them against all the faces of the people who have been previously taught to the device.

II. LITERATURE SURVEY

Many researchers have contributed to the current field. Varied mixtures of existing technologies are used. Braille systems, screen magnifiers, etc. went through some developments however later sweet-faced technical problems.

Pilling et al. conducted a study to work out whether or not the web provides opportunities for disabled folks to hold out activities that they were antecedently unable to try and do or whether or not it results in larger social exclusion.

Sinks and kings et al. states that there's no noted analysis to work out the explanations folks with disabilities can't access the web. Muller et al on the opposite hand state that the first barrier to unavailability is that of economic and technical capabilities.

This thought is seconded by Kirsty et al. World Health Organization states that dangerous HTML code and use of pdf causes a hindrance in accessing the web for the visually impaired though the W3C mentions a listing of pointers for maintaining a high level of accessibility for the visually impaired.

Power et al, states that solely fifty.4% of the issues encountered by users were lined by Success Criteria within the website Accessibility pointers a pair of.0 (WCAG 2.0) and 16.7% of internet sites enforced techniques suggested in WCAG a pair of.0 however the techniques didn't solve the issues.

Android phone-controlled voice gesture and bit screen operated chair wherever voice and gesture are recognized through golem. Developers additionally created a universal voice management on golem that is employed to launch golem applications via voice commands.

For communication the system's output to the user further as for confirming the user input, the recognized input is complete back to the user exploitation the Python text-to-speech library (pyttsx3).

The modules square measure written in Python and build use of chemical element for automation of the various module and delightful Soup for scraping the contents of the online page. The "Script" part of every module consists of custom code that entails the options of the web site contained within the module.

For instance, the Wikipedia module consists of a matter and Answer and outline feature in conjunction with the standard feature of reading out the complete article. The previous is enforced by coaching a BERT model on the Stanford Question responsive Dataset (SQuAD).

The arthropod genus that holds the system along is written in Flask. The software package is working system freelance to support hassle-free application and usage of the system.

III. SYSTEM ARCHITECTURE

The system includes a standard client-server distributed design. The system consists of the most menu that initial runs on the startup of the computer code and therefore the web site modules. The shopper communicates with the server and back with the utilization of REST Apis, therefore the web site modules aren't native to the shopper. Throughout the system, the user communicates with the software code via a speech-to-text interface. The Google library of speech-to-text (Speech Recognition) for Python is employed for this purpose.

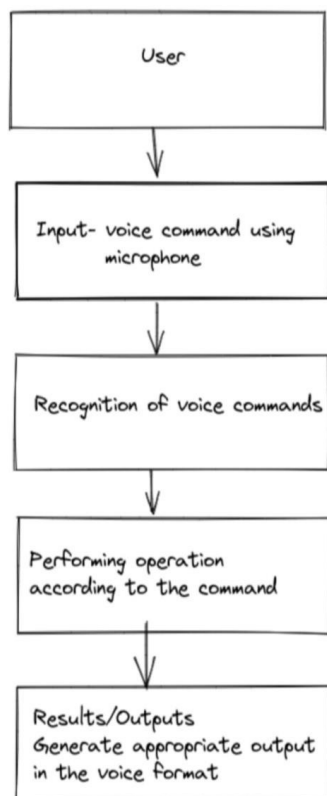
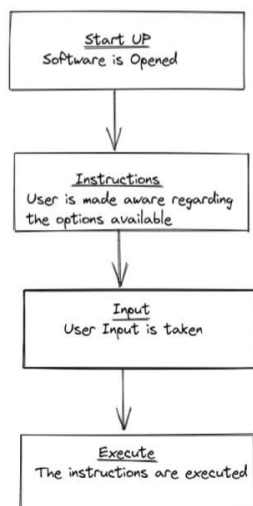


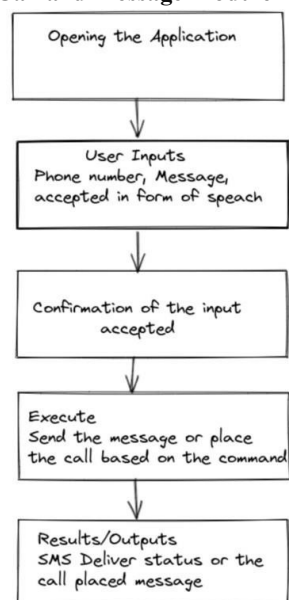
Fig: System Architecture

IV. METHODOLOGY

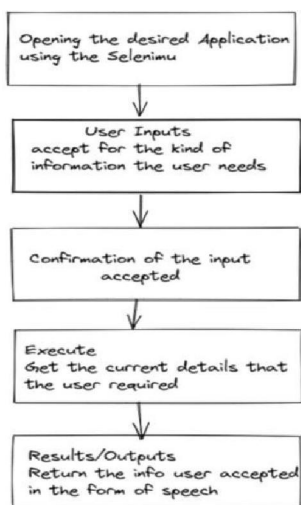
Main Menu



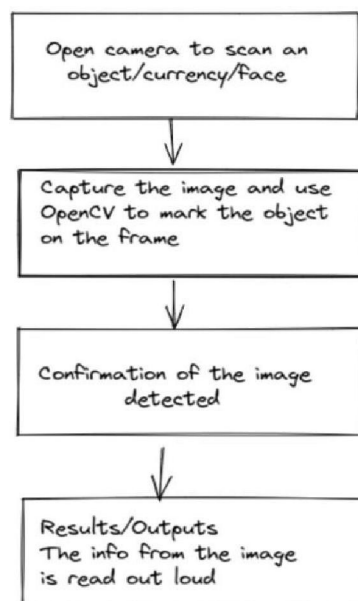
Call and Message Module



Time/Date/Battery Level



Currency Detection



V. PROPOSED SYSTEM

The major challenge in developing stable software is to involve a few keystrokes as possible and to provide an end to end experience with the help of voice. The indulgence of multiple languages and setting the right pace of the speech when played back to the user are important factors to consider.

To support the widespread usage of the software, a important parameter is the dependency of the software on the local environment and operating systems. While the tech has evolved immensely, the accessibility, especially the internet for the special cases is still stagnant.

The first main objective is to bridge the accessibility gap between the common user and also the visually impaired people with regards to the web.

The internet is blind to the visually impaired, however to not build the converse the reality, in here gift associate degree end-to-end voice-based software system for the visually impaired to modify them to access the web with stripped to no keystrokes needed.

The user can offer the commands he desires to execute as a voice input rather than employing a keyboard. The software system then uses a speech to text module to convert the input speech to text which can be the command to be executed. The command is executed using selenium web driver. Once executed the user will have three options: - either to read the entire content of the website, read a summary or ask a question.

The second and third options are implemented using machine learning. Once the voice input is taken and the command is executed the output is said to user using the text to speech module.

Thus, the software manages to make the internet more accessible easily, quickly and more effectively for the visually impaired.

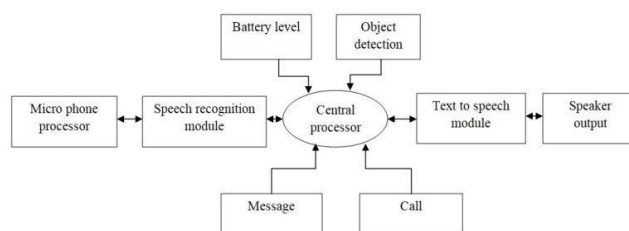


Fig: Proposed System

VI. IMPLEMENTATION

6.1 Currency Detection

User will specifically request for currency information to that the AI can respond providing the currency information. It consists of two main actors that's the user and AI. AI can reply to all user connected queries. Application can capture the image and therefore the images are going to be processed by AI and supply appropriate response like the currency detected and their kind

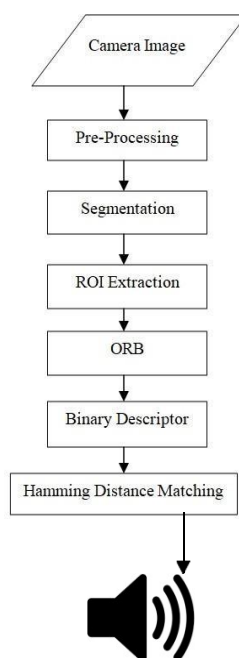


Fig: Currency Recognition

6.2 Working

In order to interact with the application, the user has to launch app. The captured image knowledge are sent to service layer wherever all the process is performed and therefore the currency is detected. A response are sent from the service layer and therefore the output are displayed within the read or the appliance screen.

User can even request for manual object detection if needed. at the side of Image process, Speech synthesis is provided to give additional accessibility to the user.

The response from the application is in the form of speech, all the detected currency related information will be informed to the user using speech to assist the visually impaired.

VII. OUTPUT

In the next section we can see how the application is able to detect the currency with the help of the camera. When the application is in use, the camera is used as background service to assist user by continuously processing surrounding data and providing information to the user. The response time of the application is less in making it useful in a real world scenario.

VIII. RESULT

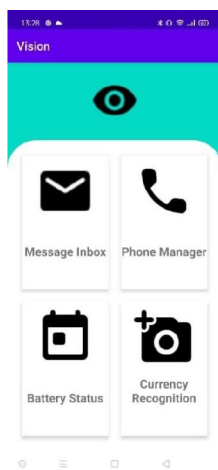


Fig: Home Page



Fig: Battery Status



Fig: Currency Recognition

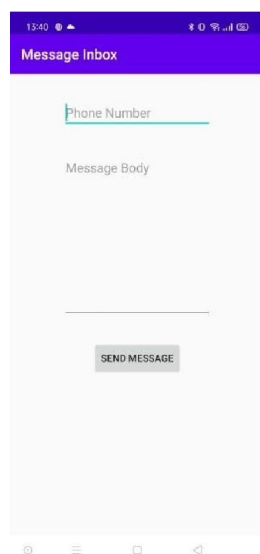


Fig: Message Inbox



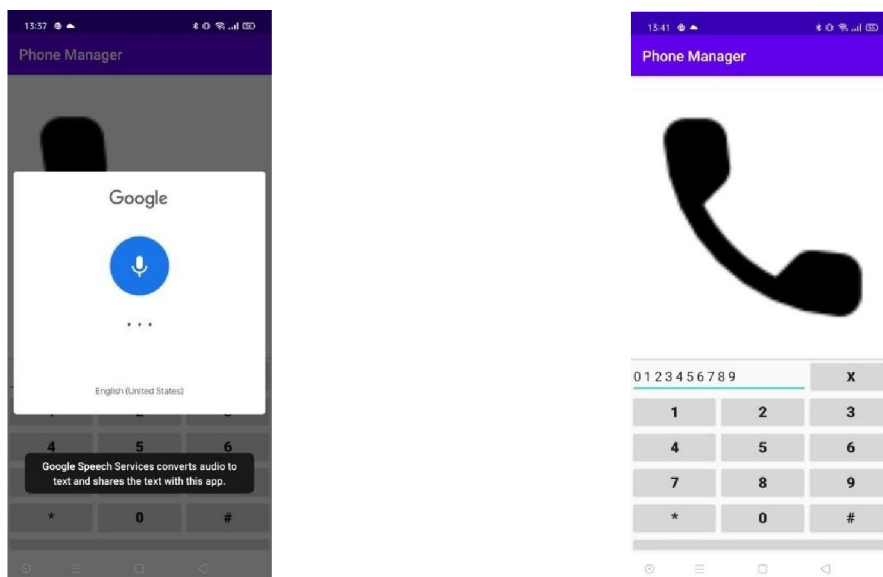


Fig: Phone Manager

IX. MODULES

- Time
- Date
- Battery Level
- Currency Recognition
- Message
- Call

X. APPLICATIONS

Virtual Assistant provides the feature of providing answers to a particular question from a given text of data, thus now the user does not have to read the entire text to figure out the answer, he/she has to simply input the question, the software will find out the answer from the text data on itself using machine learning. The software also provides a summary of the text using machine learning, so the user doesn't have to read the entire thing and thus making it easy to access the website. Thus, using machine learning and speech to text techniques we make the task of accessing the website, which was earlier difficult Now super easy, quick and efficient. Thus, we believe that virtual assistants for the visually impaired are the beginning of Web 3.0.

XI. CONCLUSION

Various techniques to implement the aforementioned system are analyzed and summarized. Various systems have different ways of implementation along with some limitations and restrictions. These types of systems are very critical for multiple reasons and the occurrence of an error in such a system/device may cause catastrophic damage and loss. The system we are achieving overcomes the limitations of the already implemented systems. Our system consists of a basic UI on a web-based application and comprises several Deep learning models; some of them are object detection, face recognition-using TF, TTS, speech recognition and so on. These modules will work together and assist in vital activities like object detection as well as face detection and recognition for the visually impaired

REFERENCES

- [1]. D. Pilling, P. Barrett and M. Floyd, Disabled people and the Internet: experiences barriers and opportunities, 2004. [online] Available: <https://www.freedomscientific.com/products/software/jaws/>.

- [2]. S. Sinks and J. King, "Adults with disabilities: Perceived barriers that prevent Internet access", CSUN 1998 Conference, 1998, March.
- [3]. Kirsty Williamson, Steve Wright, Don Schauder and Amanda Bow, "The internet for the blind and visually impaired", Journal of Computer-Mediated Communication, vol. 7, no. 1, October 2001.
- [4]. Xiangrong Chen and A.L. Yuille, "Detecting and reading text in natural scenes" in Washington DC USA IEEE Xplore, Los Angeles, CA, USA:Dept. of Stat., California Univ., 2004.
- [5]. Vivek Pradeep, Gerard Medioni and James Weinland, "Robot vision for the visually impaired" in San Francisco CA USA IEEE Xplore, Los Angeles, USA:University of Southern California, 2010.
- [6]. Rahaman, "Development of an intelligent drowsiness detection system for drivers using image processing technique", 2020 IEEE Student Conference on Research and Development (SCORED), 27-29 Sep. 2020.
- [7]. "SSD: Single Shot Detector for object detection using MultiBox", 30 Nov. 2019, [Online] Available: <https://towardsdatascience.com/>.
- [8]. Ashwani Kumar, Zuopeng Justin Zhang & Hongbo Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm", EURASIP Journal on Wireless Communications and Networking (2020), SpringerOpen, 17 October 2020.
- [9]. Sakshi Gupta, Dr. T. Uma Devi, "YOLOv2 based Real Time Object Detection", International Journal of Computer Science Trends and Technology (IJCSCT)- Volume 8 Issue 3, May-June 2020.
- [10]. R. Piyare and S. R. Lee, "Smart Home-Control and Monitoring System Using SmartPhone", 1st International Conference on Convergence and its Application (ICCA), 2013.
- [11]. P. Milhorat, S. Schogl, G. Chollet, J. Boudy, A. Esposito and G. Pelosi, "Building the next generation of Personal Digital Assistants", 1st International Conference on Advanced Technologies for Signal and Image Processing, 2014.
- [12]. J. G. Linvill, "RESEARCH AND DEVELOPMENT OF TACTILE FACSIMILE READING AID FOR THE BLIND (The Optacon)", Stanford California, 1973.
- [13]. "Cloud Text To Speech Basics", Google, July 2019, [online] Available: <https://cloud.google.com/speech-to-text/docs/basics>.
- [14]. Google Cloud, Cloud Text-to-Speech basics, July 2019, [online] Available: <https://cloud.google.com/text-to-speech/docs/basics>.
- [15]. A. M. D. Celebre, A. Z. D. Dubouzet, I. B. A. Medina, A. N. M. Surposa and R. C. Gustilo, "Home Automation Using Raspberry Pi through Siri", 8th IEEE International Conference Humanoid Nanotechnology Information Technology Communication and Control Environment and Management (HNICEM) Cebu City, 2015.
- [16]. Mishra, P. Makula, A. Kumar, K. Karan and V. K. Mittal, "A Voice-Controlled Personal Assistant Robot", 2015 International Conference on Industrial Instrumentation and Control (ICIC), 2015.
- [17]. Piyush Vashista, Juginder Pal Singh, Pranav Jain and Jitendra Kumar, "Raspberry PI based voice-operated personal assistant", International Conference on Electronics and Communication and Aerospace Technology (ICECA), 2019.
- [18]. Rui Liu, Zhihua Yu, Daili Mo, Yunfei Cai, "An Improved Faster-RCNN Algorithm for Object Detection in Remote Sensing Images", 2020 39th Chinese Control Conference (CCC), IEEE, 27-29 July 2020.
- [19]. Ma Songhui, Shi Mingming, Hu Chufeng, "Objects detection and location based on mask RCNN and stereo vision", 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 1-3 November 2019.
- [20]. Amin Azizi Suhaiman, Zazilah May, Noor A'in A.