Assistant for Visually Impaired

Sushama Khanvilkar,Bhakti Sharma,Lenora Silveira, Devika Vengurlekar,

[1]Assistant Professor, [2,3,4]Student
Computer Department, Xavier Institute of Engineering, Mahim, Mumbai, India.

ABSTRACT

Visually impaired people face a lot of difficulties in their daily life and many times they rely on others for help. Several technologies for the assistance of the visually impaired people have been developed. Among the various technologies being utilized to assist the visually impaired people, Computer Vision based solutions are emerging as one of the most promising options due to their affordability and accessibility. The proposed system created is an application for the visually impaired people to detect the objects in their surroundings and to read texts from books, newspapers, etc. This system is constructed so as to reduce the troubles and tribulations faced by the visually impaired people on a day-to-day basis. The application is constructed to bring attention to the trouble faced by these people on a daily basis while performing trivial tasks such as reading labels, navigating through objects in the surroundings, etc. The system is created to provide ease to the visually impaired people and thereby reduce the problems faced by them on a daily basis.

Keywords:-*Object detection, Visual Impairment, Navigation.*

1. INTRODUCTION

Globally the number of people of all ages visually impaired is estimated to be 285 million, of whom 39 million are blind. With the recent advances in inclusive technology, it is possible to extend the support given to people with visual impairment. Our aim is to create a user friendly system and vision substitution that uses auditory perception to gain accuracy. The functionality of the system addresses the identification of the objects and reading labels. As visually impaired people face tribulations on a daily basis, the system is constructed by helping them access the various features which enhance their development. The system is created in such a manner that the environment is easily accessible to them. This mobile application for the visually impaired people is built using Microsoft Azure and React Native.

The estimated number of individuals who are visually impaired in the world is approximately 300 million, 54 million blind, and 246 million having low vision. In today's advanced high-tech world, the need for independent living is recognized within the case of visually impaired people who face a significant downside of social limitation. They are a vital part of our society and that we should embrace them within the functioning of our society. Make an environment that's easily accessible to them. It is very difficult for them to perceive the outside world. This project aims to assist visually impaired people to own access to most of the essential features enhancing their living conditions. The system uses technologies like machine learning, image processing.

2. EXISTING SYSTEMS

Many researchers have contributed to this field. Various combinations of existing technologies have been used. The advancements in the technological fields of Computer Vision, Object detection and text-to-speech have enabled the progress in the direction of applications to make the lives of those who need it easier. These technologies can be used to sort classified images, have voice aided driving or navigation systems like SLAM systems etc. This system is specially designed to assist the visually impaired person. It facilitates in communicating the messages to the user. This enables the visually impaired person to walk independently through the most suitable navigation system.

Mobile applications are constructed for the visually impaired people. This is portable and becomes more convenient for the visually impaired people to access the various features on the application.Also real time detection systems are implemented using a combination of edge cues color histograms to be able to detect on ground obstacles and giving feedback to the user through vibration and voice on command.

3. PROPOSED SYSTEM

Computer Vision was an important part of our solution since the output of this module was received after analyzing the image captured. Computer Vision is a newly advancing technology that includes methods for acquiring the image, along with processing, analyzing and ending with understanding digital images, while performing extraction of high dimensional data from the real world in order to produce numerical or symbolic information. It allows the computer to find features more accurately and efficiently than a human would. CV models require huge tremendous resources and are tricky to implement correctly. It enables developers to develop apps with little knowledge about Machine Learning. The Azure Cognitive Services we have used are implemented as the REST API's. The main goal of this project is to provide the user with a platform which encompasses concepts of different custom layouts using Machine Learning, Artificial Intelligence, etc. We have used two features for the application. A user can recognize objects and can hear any text on any document or photo through voice. So, for these two features we have used the Cognitive Services of Microsoft Azure which helps to detect any object and to recognize text.

We have implemented a describe function wherein the system describes the surroundings being captured by the mobile camera. The object recognition is implemented wherein the platform is used for detection of an object that is in front and based on the surrounding after which the data will be sent to the algorithm for processing. After that it will process a response data, where it will translate the information into a voice through a speaker to inform the user either to go to the right side or to the left side. Another feature is the text to speech in which the system reads what is written in a document and reads it aloud.

3.1 Object Detection

In this case, when a user wants to detect what object is it or what is the name of a particular object the user needs to open the camera. For implementing the object detection feature, we used the mobile camera through Expo on the phoneits can detect camera hardware information and eventually open the camera. Vision-based object detection provides object behavior information of objects and is an intuitive detection method similar to human visual perception For detecting object input is taken by using a camera and feeds the input to the API then eventually makes the prediction. The object detection feature of the Analyze API has certain constraints. It is necessary to note the presence of false negatives due to some limitations of object detection namely small objects are not detected, closely organized objects can be difficult to identify. This functionality is used to process the relationships between the objects in an image. It also determines if there are multiple instances of the same tag in an image.

3.2 Text Recognition

The READ API is a part of Microsoft's Cognitive Services- Computer Vision . It is the OCR technology that helps extract important information to be conveyed to the user like printed text (in several languages), handwritten text (English only), digits, and currency symbols from images and multi-page PDF documents. It supports detecting both printed and handwritten text in the same image or document.OCR gives a machine power to read any text. OCR is the best way to recognize any text which makes visually impaired users able to hear text out loud.

Camera takes individual frames using detect methods and TextRecognizer. To read any text straight from camera, when any texts become available we have implemented a processor For this, Read is used to detect image and process the image that determines what text appears within the image and finally make an audio synthetic voice .The Read API is optimized for text-heavy images and multi-page, mixed language, and mixed type documents. The Read operation executes asynchronously. When you call the Read operation, the call returns with a response header called 'Operation-Location'. The 'Operation-Location' header contains a URL with the Operation Id to be used in the second step. In the second step, you use the Get Read Result operation to fetch the detected text lines and words as part of the JSON response. The time for completion of the text extraction process depends on the volume of the text and the number of pages in the document.

4. System Analysis

In the previous sections, we have defined our problems clearly and surveyed the existing literature and approaches. Now, the problem is analyzed from our perspective. We mapped how we can combine all the requirements in one system to help the target audience. The users need a system that will inform them of their surroundings, read for them the text being considered, provide the users with miscellaneous daily updates etc. The surrounding can be explained through image analysis and its description. Processing the region in front of the camera and processing it to extract features and information and supply it to the user through human-like audio. The visualization of objects with the help of sound helps to understand the view better. Next main requirement is converting the text to speech as the users cannot read this will be done with the help of CV and the text is extracted with OCR for better accuracy. Besides being an assistant the hidden necessity is convenience for which this system is developed into an application which can be used on smartphones.

4.1 Use Case Diagram

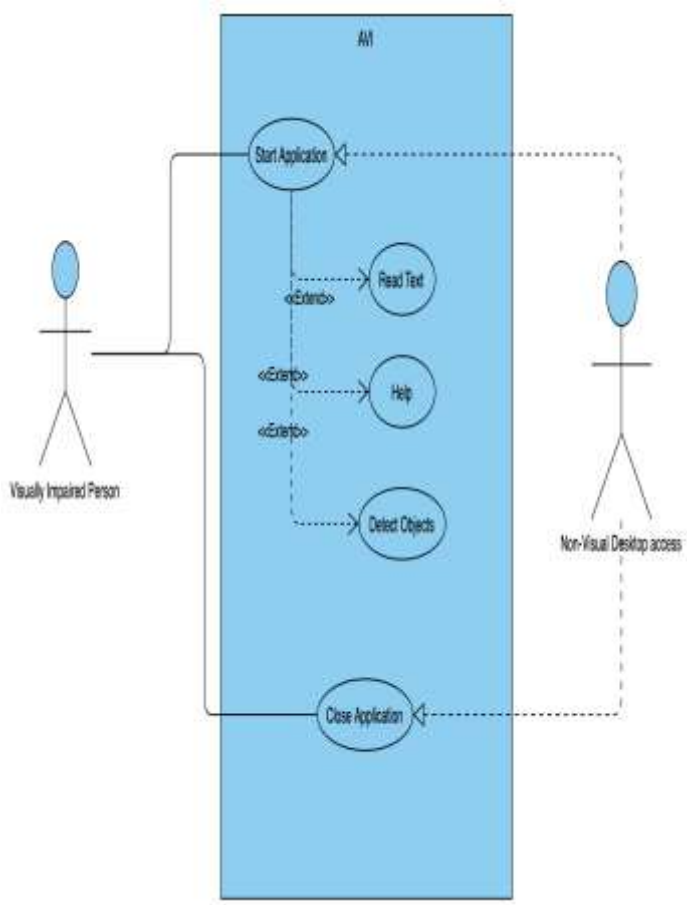

Figure 3.1: Use Case Diagram of our Proposed System

4.2 GUI Design

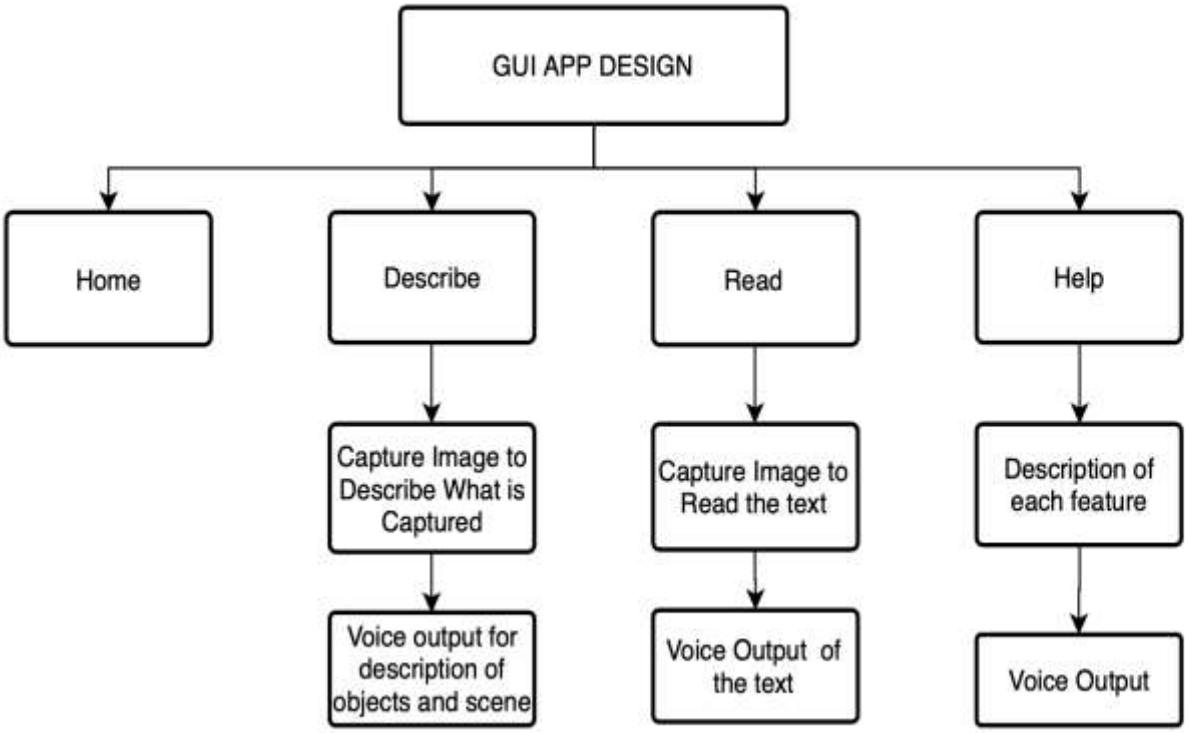

Figure 3.2. GUI Diagram

4.3 Flowchart

The high-level steps involved during the various phases of the project using flowcharts is explained in this section.

4.3.1 Flowchart detailing the Object Detection process

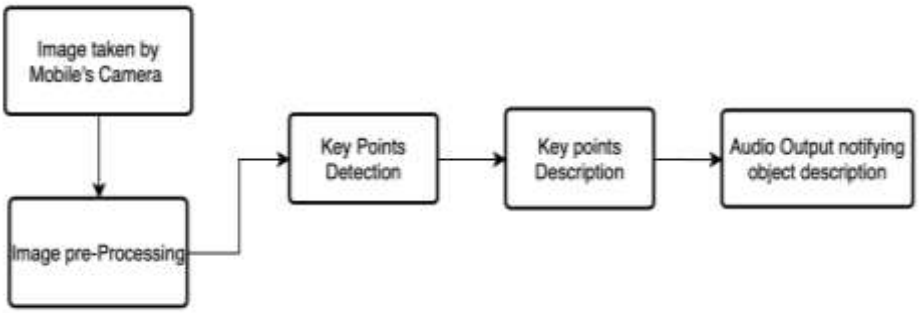

Figure 4.3.1: Flowchart explaining the process for Object Detection

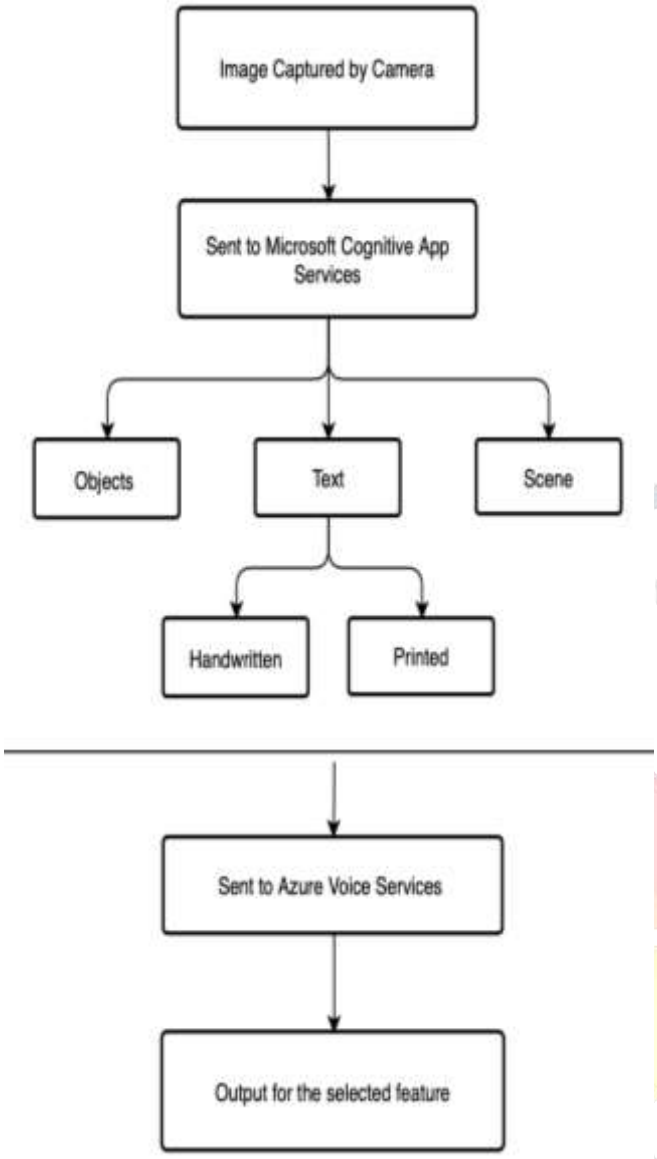

Figure 4.4: Flowchart for proposed system

4.4 Image analysis in computer vision.

4.4.1 Computer Vision

The Computer Vision API we used is a platform to process images and get information related to it. Once the image or image url is given through the API it analyzes the content in different methods based on the input choice of the user. This API proves as a baseline for our assistant as it is in parallel to the objectives of our project. It boosts the content discoverability which is crucial for the execution of our system. It uses processing of visual data to label the objects , extract text and moderate content being captured from the camera.

4.4.2 Analyze API

This API has many features out of which we used the following for execution of the required functionalities. In a perfect world, people would not upload inappropriate content to apps. The Azure Computer Vision can detect and score many features in the real world.

Detecting Objects

Here the API returns tags applied to the object. If an image contains a tree, a dog and a car the Detect operation returns all those tags with their coordinates of positions in the image. There is no relation between tagging and object detection taxonomy. This is further used to define relationships of objects with each other in the respective image. It specifies if there is the presence of the same tag more than once in the same image. The Detect API identifies only the objects and living beings but with TAG API an extra description of the objects being indoor or outdoor is also provided.

Describing an image

The function of this feature is the most essential in our project as it describes the whole image for the user. This is important for a visually impaired person to know about their surroundings. It is used for them to be aware of what is being captured through their app lens. The Computer Visions algorithms generate various descriptions as an output for the objects present in the image. It then generates a human language sentence for the description. The confidence score from each description is then evaluated and a list of the highest to lowest confidence score is generated.

5. Implementation Methodology

5.1 Proposed Methodology

In this application model, the Computer Vision API developed by Microsoft is used. It will enable the environment description and text reading from images within the application. In addition, we have integrated the Analyze API, Read API two such cognitive services provided by Microsoft Azure Cloud platform. These are easily accessible for the developers.

AVI is a mobile application that is completely voice based making it easy and simple to navigate. Responding to the user with a human voice is the output of the system. AVI mobile application is built using React Native and Microsoft Azure.

The application can be divided into two standard parts

- Front end
- Back End

5.1.1 Frontend

The frontend consists of a mobile application built with React Native which is a Javascript framework . We used React to learn and explore a framework under the domain of reusable components. React provides a platform that makes the fusion between JavaScript and HTML. It is preferable as it helps manage and maintain states across components. We also used Expo for making it a universal native app for cross-platform usage and considering all types of users. Because the users are visually impaired, this application interface is designed to meet the necessary requirements. As making the app uncomplicated was a priority the layout of the app was kept simple and easy to navigate. The icons for the features 'Describe' and 'Read' were made big and visible. Voice commands were also implemented to maneuver through the app. Along the features we also provided a help section for the user to depend on whenever required.

5.1.2 Backend

It is made of FastAPI framework that serves the requests sent from the mobile application. We made use of this framework as it supports asynchronicity and machine learning models that are accessed by REST API's. The server was hosted on Azure App Services that completely managed the web hosting service for building the RESTful API's. This being provided made it easier to focus on the prioritized features rather than worrying about infrastructure maintenance, security patching, and scaling. Depending on what event has been requested the server communicates with the Azure services to respond back with the desired output. The images which are captured by the camera are automatically sent to the processor through Wi-Fi, which analyses the objects present in the images, identifies them

and gathers all the information about the captured image in collaboration with the different API's. The information about the objects in the captured image that are processed by the API, is converted into the audio file. The audio output is sent to the user which can be heard through the smart phone's speakers or earphones if connected.

5.1.3.API Implementation

Azure's Computer Vision service gives access to advanced algorithms that process images and return information based on the visual features you're interested in. To access the services of Azure and connect the application with the required API's of the Computer Vision service the key and endpoint from the resource has to be added to the code in the folders of the project. Also the link received from the services after running ngrok has to be pasted in the config folder.

5.1.4 Features

The app composes of two main functionalities:

5.1.4.1 Describe

This feature is for the description of the region captured by the camera at a given instance. It not only detects objects but also describes the surrounding area. We observed that many existing applications make use of Object Descriptors that are trained on the COCO dataset which compared to Microsoft Azure Cognitive Service limits the number of objects being identified by the app. Hence, we moved further with Cognitive services. The output that is received after the computation is in the form of tags allotted to objects that are identified upon processing it. These tags are further clubbed into human perceived language to make better sense and given as output in the form of audio that is read out by the Expo component of React. Cycles, cars or pedestrians. The vision description service can start when users ask about what things are in front of them self. The system can automatically capture the image of their foreground and send it to the back-end server provided by us to compute for the result

5.1.4.2 Read

There are many public places like banks, restaurants, malls, etc that do not have the facility for those with visual impairment to read information written on menus, boards and signs. This leads to the text being inaccessible to them. To help with this issue we used Microsoft Azure Cognitive Service - Read which is optimized for analyzing images with heavy text, mixed language and different types of documents. The extracted text is converted to speech for the user to understand what is written and the function to serve its main purpose.

6. Details of Hardware and Software

6.1 Details of Software

1. Azure Cognitive Services

Cognitive Services brings AI within reach of every developer—without requiring machine-learning expertise. All it takes is an API call to embed the ability to see, hear, speak, search, understand and accelerate decision-making into your apps. Enable developers of all skill levels to easily add AI capabilities to their apps.
Benefits of Cognitive Services

- Apply AI to more scenarios with the most comprehensive portfolio of domain-specific AI capabilities on the market.
- Build confidently with the first AI services to achieve human parity in computer vision, speech and language.
- Deploy Cognitive Services anywhere from the cloud to the edge with containers.
- Get started quickly—no machine-learning expertise required.

2. Azure App Services

A fully managed platform for building, deploying and scaling your web apps quickly build, deploy and scale web apps and APIs on your terms. Work with .NET, .NET Core, Node.js, Java, Python or PHP, in containers or running on Windows or Linux. Meet rigorous, enterprise-grade performance, security and compliance requirements using a trusted, fully managed platform that handles over 40 billion requests per day.

Features of Azure App Services:

- Fully managed platform with built-in infrastructure maintenance, security patching and scaling.
- Built-in CI/CD integration and zero-downtime deployments.
- Integration with virtual networks and ability to run in an isolated and dedicated App Service environment.
- Rigorous security and compliance, including SOC and PCI, for seamless deployments across public cloud, Azure Government and on-premises environments.

3. FastAPI

FastAPI is a modern, fast, web framework for building APIs with Python 3.6+ based on standard Python type hints. Independent TechEmpower benchmarks show FastAPI applications running under Uvicorn as one of the fastest Python frameworks available, only below Starlette and Uvicorn themselves.

- uvicorn - for the server that loads and serves your application.

The key features are:

- **Fast:** Very high performance, on par with NodeJS and Go (thanks to Starlette and Pydantic). One of the fastest Python frameworks available.
- **Fast to code:** Increase the speed to develop features by about 200% to 300%.
- **Fewer bugs:** Reduce about 40% of human (developer) induced errors.
- **Intuitive:** Great editor support. Completion everywhere. Less time debugging.
- **Easy:** Designed to be easy to use and learn. Less time reading docs.
- **Short:** Minimize code duplication. Multiple features from each parameter declaration. Fewer bugs.
- **Robust:** Get production-ready code. With automatic interactive documentation.

4. React Native

React Native is an open source framework for building Android and iOS applications using React and the app platform's native capabilities. With React Native, you use JavaScript to access your platform's APIs as well as to describe the appearance and behavior of your UI using React components: bundles of reusable, nestable code. Many different kinds of people use React Native: from advanced iOS developers to React beginners, to people getting started programming for the first time in their career. To work with React Native, we will need to have an understanding of JavaScript fundamentals.

6.2 Hardware Details

- Smartphone with a camera that supports ExpoGo application
- Computer System with 8GB RAM minimum
- Speed of processor 2.5Ghz

7. OUTPUT

The output of the application is an audio that
is generated after performing the describe and the read functions. The input is taken by the camera for both the functions. While performing the describe function, the input taken by the camera is the objects that are present in the vicinity of the visually impaired people and the respective audio output is

generated specifying the name of the objects detected by the camera so as to help these people identify the objects easily. While executing the read function, the inputs taken are the texts from a book, cover of a book, etc. An audio is generated as an output after the texts are detected by the camera and are heard by the visually impaired people. Both the describe and the read functions help the visually impaired people to reduce the tribulations faced by them on the daily basis and thereby lead to great convenience and ease to perform the tasks.

8. CONCLUSION

The research for helping those who need assistance has been growing rapidly in the past few years with the development in AI and machine learning along with cloud services. The project as a whole cannot gain any profit from the app as it is a non-profit product. However the cause itself is encouraging investors to fund the researchers in this motive. The main idea behind this project is to help the visually impaired people to detect the objects around them with ease and also helps them to hear audio of the read texts from the paper, cover of the book or anywhere else. It aims to provide a working environment in which all objects can be accessed smartly, giving them access to information and technologies particularly. The well documented functionalities made it easier to build the system to assist the visually impaired people facing social restrictiveness to perform daily tasks.

9. FUTURE SCOPE

We have developed the application considering the visually impaired and to help them experience the world better. This project focuses on describing the surroundings with the range of the camera lens and reading the text in front and converting it to speech. There is a wide range of scope for this project, and it can be enhanced to interpret and analyze images and text with much more context. Our system can be deployed with Iot appropriately.

With further using more of the available Azure's Cognitive Services, there can be more functionalities added that will be of help:

- Face Emotion Recognizer: Using this, the visually impaired can identify the emotions that the people sitting right in front of them are feeling.
- Product Scanner: Using which they can find if a particular product is fresh or is past its expiry date.
- Medicine Reader: This not only will help them read the medicine names but also provide information related to the medicine.
- Support for multiple languages: we plan to introduce multi-lingual support for voice commands and speech. This would make it accessible to a larger number of people.
- Enhancement to this system can be done by adding the features of currency recognition.

10. ACKNOWLEDGEMENT

We would like to express our deep and sincere gratitude to our Prof. Sushama Khanvilkar who gave us the golden opportunity to do the project on the topic entitled Assistant for Visually Impaired People. It has helped us in doing a lot of research for learning new concepts and thereby executing the tasks in a perfect way.

11. REFERENCES

1. Vyavahare, P., & Habeeb, S. (2018). Assistant for Visually Impaired using Computer Vision. 2018 1st International Conference on Advanced Research in Engineering Sciences (ARES).
2. Sultan, M. R., & Hoque, M. M. (2019). ABYS(Always By Your Side): A Virtual Assistant for Visually Impaired Persons. 2019 22nd International Conference on Computer and Information Technology (ICCIT).
3. Khan Shishir, M. A., Rashid Fahim, S., Habib, F. M., & Farah, T. (2019). Eye Assistant : Using mobile application to help the visually impaired. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT).
4. Chen, R., Tian, Z., Liu, H., Zhao, F., Zhang, S., & Liu, H. (2018). Construction of a voice driven life assistant system for visually impaired people. 2018 International Conference on Artificial Intelligence and Big Data
5. Shubham Melvin Felix, Sumer Kumar, A. Veeramuthu "A Smart Personal AI Assistant for Visually Impaired People", 2018.
6. Microsoft-Azure Cognitive Services documentation

.