Recent Trends in Intensive Computing
M. Rajesh et al. (Eds.)

doi:10.3233/APC210204

VirtualEye: Android Application for the Visually Impaired

Jay Bagrecha [a, 1], Tanay Shah [a], Karan Shah [a], Tanvi Gandhi [a], and Sushila Palwe [a]

[a] *School of Computer Engineering & Technology, MIT World University, Pune, Maharashtra, India*

Abstract. In India, almost 18 million visually impaired people have difficulties in managing their day-to-day activities. Hence, there is a need to develop an application that can assist them every time and give vocal instructions in both English and Hindi. In this paper, we introduced a robust lightweight Android application that facilitates visually impaired individuals by providing a variety of essential features such as object and distance detection, Indian currency note detection, and optical character recognition that can enhance their quality of life. This application aims to have a user-friendly GUI well suited to the needs of the blind user and modules like Object Recognition with Image Captioning so that the visually challenged user can gain a better understanding of their surroundings.

Keywords. Mobile Application; Visually Impaired; Object Detection; Image Captioning; Indian Currency Recognition; Optical Character Recognition; Deep learning; text-to-speech.

1. Introduction

Globally there are around 280 million people, who are visually impaired, of whom about 40 million people suffer from complete blindness. Considering India's vast population, it alone has 12 million people who live from blindness. With the rapid advancement in technology, many mobile devices include applications and features to help visually impaired people but most of the devices are designed for people with vision. One of the most challenging tasks for a visually impaired person is to identify the day-to-day things around them. With the availability of mobile devices and the rising computational capabilities, these people can be assisted using artificial intelligence and computer vision techniques. Imagine being a visually impaired person, able to locate and track the simplest of things like chairs, bookshelves, cupboards, and all sorts of daily items with ease. Today's technology has made life better for all. Artificial Intelligence has become entrenched in our everyday routines. As a result,there was a need to build an application that could assist visually disabled people with day-to-day activities using AI.

[1] Jay Bagrecha, School of Computer Engineering & Technology, MIT World University
Email: jaybagrecha1234@gmail.com.

2. Literature Review

2.1. ***Object detection with Image captioning:***

According to Faizad Amin in [1], If we want to understand the geometry of a scene, it is important to identify the obstacle present in our way as well as estimate the depth. There are two types of Object Detection: they can be static or dynamic. It is easy to locate static objects as they are fixed but it is a bit difficult for dynamic objects as they are constantly moving. In this approach, stereo images are used for object recognition. We calculate the unevenness of the detected objects and then calculate the distance of these objects from the camera which we call "depth". In Real-Time Object Detection Application by Selman Tosun [2], the visually weakened people will be able to recognize the obstacles while they are walking on the road using the feedback which they will get in form of audio and this will help them prevent possible accidents. The operations are performed using the inbuilt audio and the camera modules. This application has different modes for both indoor and outdoor transportation, voice feedback is a plus. According to Xiaofei Fu [3] in his Mobile Application for visually impaired people, they have made it possible to create most of the functionalities of the app offline. The functionalities include face detection, gender classification. They have made a system where the contents of the picture are given as output through voice. In the Intelligent Eye application by M Awad [11], there are features like the detection of objects, banknote, light, as well as color. All these features work completely fine even when the device is offline. The accuracy is also very good in this application and here the focus is on some different features like light and color detection, which are useful from the point of view of blind people. As we saw in Xiaofei Fu's Mobile Application [3], there is a provision of getting output through voice, so Quan Zheng You's [4] approach gives a perfect explanation of the current scene, instead of just naming the objects. Here an automatic generated natural language description of an image is given as output. When an image is fed to the CNN, then first the extraction of the top-down visual features are done and at the same time the visual concepts like the attributes, regions are detected, and proper structuring of all these words are done and the whole sentence is given as output. According to O Vinyals [12] for Image Captioning, A RNN is used for the generation of sentences that follows CNN for encoding the image into a proper representation, and then the Probabilistic Neural network approach is used, and the softmax algorithm is used for word prediction, which is very light.

2.2. ***Indian Currency Recognition:***

The Recognition of the different denominations of Indian banknotes has been tackled using several methods and a considerable amount of research has been done in the field of currency note detection. Over the years the researchers presented their work based on characteristics like color, texture, etc., and have used many Deep Neural Networks and Machine Learning algorithms like CNN, ANN, RCNN, Masked RCNN, PCA, Naive Bayes classifiers, Random Forest, etc. for their research. According to [5] paper, a simple Convolution Neural Network architecture-based method was used to train the model and the implementation of currency recognition was executed in both web and android applications. Some of the frameworks used were TensorFlow, TensorFlowLite for android, and at last, enhanced their model by performing hyper-parameter tuning. New Dataset was created and then Data Augmentation was applied

to get 11657 images. A detailed tabular comparison based on Training and Testing accuracies, Computation Time, and overall performance of the models was done by the authors with many popular pre-trained models like VGG19, Xception, Resnet50, Alexnet, InceptionV3, and with a simple CNN model proposed in [6]. The results were quite fascinating and the proposed model outperformed all other models by giving the best training and testing accuracy of 100% and 87.5%. According to [6] paper, a DL model was used to detect the different denominations of Indian Banknotes. A Pre-trained model MobileNet was used that is a transfer learning method available in Keras Applications. The creation of a new dataset for four different denominations of the Indian currency and performing data augmentation on the dataset was done to get 12160 images. Results of their classification framework were good enough with a Training and Testing accuracy of 100% and 96.6%. The authors of this paper say that their approach requires very little data preprocessing and will perform great even if the input images have some disturbances or if the images are unclear. According to [7] paper, an android application was developed by the authors especially for blind people so that they can easily know the denominations of the Indian Banknotes. The authors of this paper have implemented a basic Deep learning model, which scans an image from your smartphone's camera and then gives an output based on some probabilities in the form of voice so that a blind person can hear it easily. The authors have gathered a dataset of all the valid Indian currencies and have performed data labeling using a labeling tool to get 2536 images. They have used another transfer learning method - Faster RCNN with Resnet v2 to get 87% accuracy and loss of 0.201.

2.3. ***Optical Character Recognition:***

According to [8] paper**,** Tesseract is a perfect engine for OCR. HP has already developed page layout analysis technology that itself was used in the products. And that's the reason why there was no need for Tesseract to have its page layout analysis**.** After the page layout analysis is done, we use the line-finding algorithm. The line-finding algorithm is used to save loss of image quality without eschewing the images. The major parts of the process include filtering blobs and construction of lines. In the first step, components are outlined and stored, this is called connected component analysis. It is computationally expensive but easier to handle white text with a black color background. Now, outlines are gathered by the process of nesting them into Blobs. Then the Blobs are simplified into text lines. Text lines are broken into words. The next stage is Recognition, which is a two-step process. In the first pass, the algorithm tries to recognize each word which is then passed to an adaptive classifier as input. Then in the second pass, unrecognized words of the page are tried to recognize again. The unique thing about Tesseract is that it handles white-on-black texts in a better way. Paper [9] talks about all the methods by which OCR is done and its major challenges. According to the [9] paper, the main phases of optical character recognition include preprocessing**.** Over here, systems utilize binary or grey images as processing color images is computationally expensive. After this, we do Segmentation. In segmentation, the system separates the text part from the input image. There are three kinds of algorithms for document segmentation: Top-down method, Bottom-up method, Hybrid method. This gives us about 98% accuracy. The next step is Normalization where the characters which were separated are reduced in size depending on the algorithm used. The image is converted in the form of m*n matrix. After this, feature extraction is performed which can be time-consuming and complex

too. OCR systems use a lot of methodologies of pattern recognition. In this, an example is assigned to each predefined class. Techniques of OCR classification are: Statistical Techniques (The main statistical methods that are performed in the area of OCR [19] are Likelihood or Bayes classifier, Nearest Neighbor (NN), Clustering Analysis, Fuzzy Set Reasoning, and Quadratic classifier, Hidden Markov Modelling (HMM)), Neural Networks, Template Matching (least complex method) and Support Vector Machine (SVM) algorithms, and Combination of the classifier. The final step is post-processing. The main objective of OCR is to decide the context of the image. OCR systems make use of a dictionary to make minor changes in the errors that the system produces.
Paper [10] discusses Cloud Vision API and CNN. The first step is to do a layout analysis to locate the text on the image. The next step is to perform a text recognition analysis to produce the text. The step is performed through a convolutional neural network. Convolutional neural networks are a subset of neural networks. CNN follows the complex structure of the Human's visual cortex that is present in the brain through which we identify objects around us. The accuracy we achieve is about 80 percent.

3. Research Gaps

On reviewing [1], [2], [3], [4] papers we figured out a few extremely important things, which are not present very widely, and we would be planning to implement and emphasize more on it while creating our project. First thing is that along with object detection it is very important to also, find the depth and the distance between the object and the user because we are trying to create an application that can be used in daily commute, so according to the depth of the object, priorities are set, for example, if in the scene if a dog and a bike both are present, but the dog is at 100 cm and the bike is at 1 m, so object detection will only mention the name of both objects, but we want that user should get first output as the dog is present at 50 cm and then about the bike. So to set priorities and to know the rough distance between the objects is very important. After reviewing [11],[12] papers we figured out that only giving the output of the name of the object and also the distance isn't sufficient, if the user is moving on the road, he would need to know about the whole scene present in front of him so that he confidently moves forward. In addition, explaining the whole scene can be done using Image Captioning, because if a person is driving a bike object detection will only give output as Bike and Person present at 200 cm, but with Image Captioning, it will give output as a person is driving a bike at 200 cm. We would also be making a system where the user has the option to get all outputs in Hindi as well because it should not happen that due to the language barrier, the user is unable to use our application. After reviewing [5], [6], [7] papers we established that the main problem of all the above-mentioned approaches is that they employ conventional pre-trained models, such as VGG16/19, AlexNet, Resnet v2, MobileNet, etc., which require a large number of annotated data. The datasets used in the papers do not have all the denominations of currency accepted in India. No previously mentioned papers incorporate the identification mechanism for counterfeit currency notes. Therefore, there is a need for another model, which is capable of extracting more deep features so that visually weak people can depend less on normal people and lead a better life. After reviewing [8], [9], [10] we established that the entire process of OCR is complex and more prone to errors if we are doing it from scratch. The software used in [9] expects us to give a processed image. So overall, there are various approaches to perform OCR. Each has its pros and

cons. Maintaining accuracy along with fast response is a challenge. There is a need to develop a system that does OCR along with Text to Speech Conversion in whichever language the user wants.

4. Proposed Methodology

4.1. ***Objectives:***

- To assist a visually impaired person about surroundings in real-time.
- To integrate Depth detection and Image Captioning with Object Detection so that they can understand what exactly is happening around them.
- To detect both static and dynamic objects present in the surrounding, and to get a proper output of those objects and the scene using NLP.
- To describe the content of an image using properly formed English sentences using Image Captioning.
- To convert formed sentences to the regional language.
- To allow the visually impaired to autonomously deal with Indian banknotes, particularly while accepting their money back during their day-to-day activities.
- To help visually impaired people know what is written on a piece of paper, or in an image, or anywhere around them, whenever needed.

4.2. ***Modules:***

This application helps visually impaired people to visualize and navigate through their surroundings. All they have to do is to launch the app using Google's Talkback feature, and then onwards, the app will take care of the rest. The android application will be built using Flutter Framework and Google Talkback utilization. User Interface will be created in Dart Language. UI will consist of Home Page, which cordially invites blind users to start the app through audio. The proposed app uses deep learning for object detection and provides the name of the object and its position relative to the user (either left side or right side) as an audio output. It promptly detects Indian Banknotes for the impaired user and also uses Optical Character Recognition to scan text and provide the price and expiry date of the product. Flutter framework was used to develop this application; thus, this app will run on Android as well as on iOS platforms. TensorFlow Lite and MobileNet provide the necessary dependencies and will do the required object detection.
A use case diagram shown in Figure 1, explains all the scenarios when a user interacts with the application.

Module 1: Object Detection with Image captioning and distance determination

In the Object Detection module, the user will just have to point the mobile's camera in the surrounding, and the app will start detecting objects in real-time. When an image or video is fed, the object recognition model can detect any set of objects in the surroundings and gives information about the position of the objects in the image. Object detection models are trained in such a way that they can analyze the location of multiple classes of objects. When we provide an image to the next model, it produces a list of identified items, the location of the bounding box containing each item, and a score indicating the correct confidence.
The model returns an array of four numbers representing the bounding rectangle that surrounds its location for each object found. Objects with maximum accuracy will be labeled and their speech output is given.

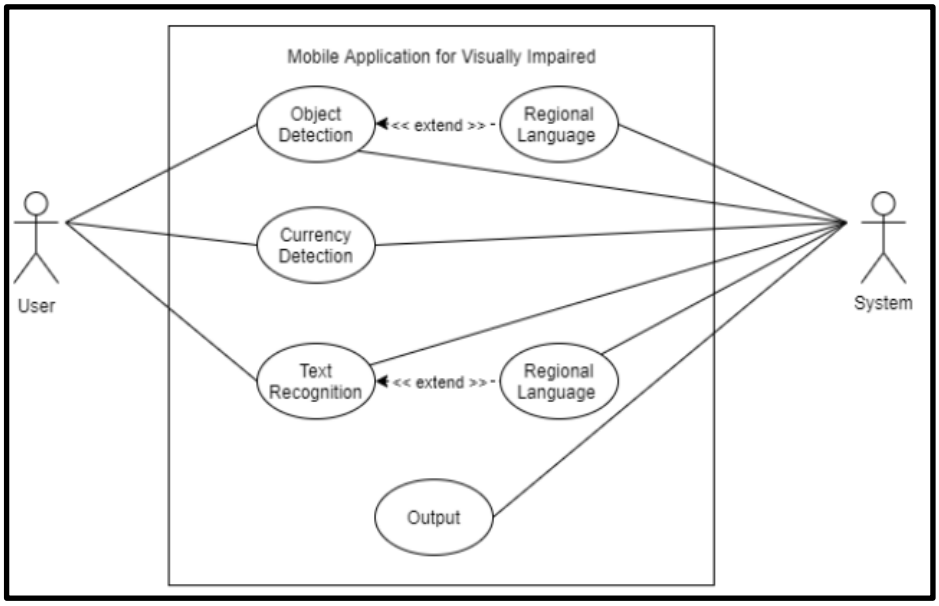


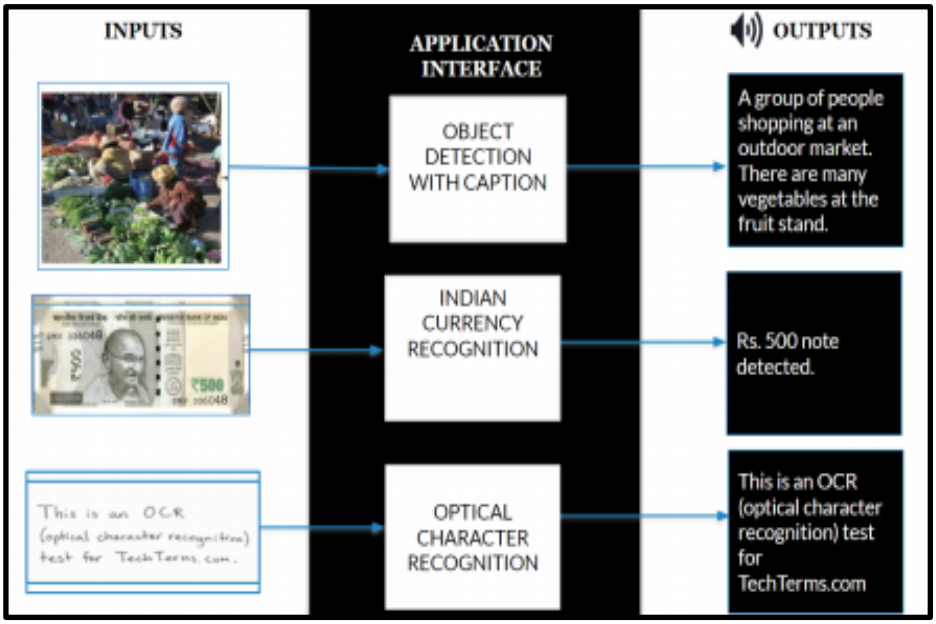
Figure 1. Use case diagram

Module 2: Indian Currency Recognition

In the Indian Banknotes recognition module, the training of a lightweight CNN model is required or any Pretrained model available in Keras Applications which not only accurately recognizes all the denomination of Indian banknotes in real-time with good accuracy but also determines that whether the banknotes given to a visually impaired individual is real or fake. This module will also concentrate on voice commands to guide the user through each move, reducing their reliance on others, particularly during outdoor activities.

Module 3: Object Character Recognition

In the OCR, the user just has to point the mobile's camera at a price tag of a product. After that, the app will tell the user about the price, expiry date, etc. of that product. OCR module is used to identify the text inside an image and separate it into actual text. Our app needs this to recognize the content on the price tags, bills, flex boards and to extract important information from various other sources. To do this we have to divide the photo into two halves. This saves extra computations. Figure 2 the lower half of a tag is where the price is written most of the time. Further, in the lower half, we then



locate the bold or the bigger size text.

Figure 2. Workflow Diagram (I/O Block Diagram)

5. Requirements

5.1. ***Hardware Requirements:***

Android Smartphone with a well-functioning camera with at least 2 GB of RAM.

5.2. ***Software Requirements:***

- TensorFlow- It is an open-source library used to implement different deep learning algorithms. With the help of TensorFlow API, a lot of mathematical and numerical computations can be done with ease. It was developed by Google such that any model can be trained and implemented easily.

- OpenCV- It's an open-source library for machine learning and computer vision applications. It was built in such a way that the use of machine perception in commercial products and provision is present for common infrastructure for computer vision applications.

- Google Cloud Vision - It is an application programming interface that enables the developer to use powerful pre-trained machine learning models including image labeling, face and landmark detection, Optical Character Recognition (OCR) through REST APIs. Using Google Cloud vision API, we can easily integrate CV with different technologies.

- Android Studio- It is an IDE for the android operating systems, built and designed specifically for Android development. It has a proper system based on Gradle.

- Python- It is an open-source general-purpose interpreted, high-level programming language used to develop web applications, software applications, games, data-science applications, and many other things.

- Flutter- It is an open-source UI software development kit developed by Google to create various applications for Windows, Android, iOS, etc from a single codebase.

- Google TalkBack - Google TalkBack is a service that gives vocal instructions to its users and allows them to access Android applications with ease, and can interact better with the device and whole Android ecosystem.

- Google Colab- Colab is a Jupyter notebook-based environment that runs purely on the cloud. It doesn't require any proper setup and different people can simultaneously work on Colab notebooks that you create the same way as we work on Google docs.

6. Conclusion

Visual impairment is one of the most debilitating disorders a man can have. It affects an individual's overall well-being as well as their emotional and social relationships. In this paper, an Android application is introduced for visually impaired individuals that assist them in visualizing and navigating their surroundings, thereby reducing their reliance on others, especially during outings. The application includes modules like object detection with image captioning and object distance from the user, Indian banknote detection that also alerts the user if the note is fake, and optical character recognition (OCR) that helps the user know what is written on a piece of paper, or in an image, or anywhere around them, whenever needed. Many different functionalities, such as barcode scanning, card readers, light identification, color recognition, road assistance, voice-based SMS and Email, Emotion Recognition, and location sharing, can be built in the future to benefit visually impaired people.

References

[1] Amin, F., Mehdi, S. A., & Khan, A. D. (2019). Distance Estimation using Tensorflow Object Detection. 2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON). doi:10.1109/upcon47278.2019.8980216

[2] TOSUN, S., & KARAARSLAN, E. (2018). Real-time object detection application for visually impaired people: Third eye. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP). doi:10.1109/idap.2018.8620773

[3] Fu, X. (n.d.). Mobile assistant app for visually impaired people, with face detection, gender classification and sound representation of image. Electrical Engineering Department, Stanford University.

[4] You, Q., Jin, H., Wang, Z., Fang, C., & Luo, J. (2016). Image captioning with semantic attention. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.503

[5] Veeramsetty, V., Singal, G., & Badal, T. (2020). Coinnet: Platform independent application to recognize Indian currency notes using deep learning techniques. Multimedia Tools and Applications, 79(31-32), 22569-22594. doi:10.1007/s11042-020-09031-0

[6] Mittal, S., & Mittal, S. (2018). Indian banknote recognition using convolutional neural network. 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU). doi:10.1109/iot-siu.2018.8519888

[7] Bhavsar, K., Jani, K., & Vanzara, R. (2020). Indian currency recognition from live video using deep learning. Communications in Computer and Information Science, 70-81. doi:10.1007/978-981-15-6648-6_6

[8] Smith, R. (2007). An overview of the tesseract ocr engine. Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2. doi:10.1109/icdar.2007.4376991

[9] Hamad, K., & Kaya, M. (2016). A detailed analysis of optical character recognition technology. International Journal of Applied Mathematics, Electronics and Computers, 4(Special Issue-1), 244-244. doi:10.18100/ijamec.270374

[10] Neves, António & Lopes, Daniel. (2016). A practical study about the Google Vision API

[11] Awad, M., Haddad, J. E., Khneisser, E., Mahmoud, T., Yaacoub, E., & Malli, M. (2018). Intelligent eye: A mobile application for assisting blind people. *2018 IEEE Middle East and North Africa Communications Conference.*doi:10.1109/menacomm.2018.8371005

[12] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2015.7298935