Re-Vision: Spectacles for the Visually Challenged

Vraj Baxi, Justin Patel, Vaibhavi Lakhani, Priyang Bhatt

Abstract: The primary issue faced by all types of visually challenged people around the globe is their self-independence. They feel dependent on every task they want to perform in their daily lives and this acts as an obstacle to the exciting things which they would otherwise want to do. This paper proposes a solution in the form of wearable smart spectacles which works on the Raspberry Pi Platform for making the visually challenged people self-sufficient and move freely in their known as well as unknown surroundings. This smart Spectacle uses a USB (Universal Serial Bus) camera and detects real-time objects in the vicinity using the SSD-MobileNets object detection algorithm and provides vision in the form of audio through the use of Headphones. The Smart Spectacles also combines the use of the OCR algorithm for Text Detection and proposes the module for quick and accurate detection of currency by the visually challenged.

Keywords: Smart Spectacles, Object Detectin, Text Detection, Currency Detection

I. INTRODUCTION

The World inhabits a huge number of visually challenged humans[1]. According to the global vision impairment facts listed by "The International Agency for the Prevention of Blindness (IAPB)", 36 million people are blind, 217 million people suffer from moderate or severe distance vision impairment and 1 billion people have near vision impairment[1]. It is estimated that the number of blind people will almost double until 2020[2]. Whether people are suffering from visual impairment or blindness, one thing which both the groups equally face is their inability to be completely independent. A visually impaired person might be able to see partially depending upon whether he/she has a mild or severe impairment whereas the situation is worse in case of a completely blind person for whom there is a total loss of vision.

From time to time various technological improvements have helped the visually challenged people perform some basic activities like reading the text. For Instance, software like screen reader has helped in reading the text present on the screen as it converts from text to speech but again for the tools to convert text to speech, the text should be in a readable format and sometimes it might be possible that the human writing is not recognized when scanned to a computer and read by the software. The existence of Braille Systems which consists of the raised dots, helps people read the letters by moving their fingers over the dots and feeling the tactile version[2]. Traditionally, Service Dogs were trained to help their owners navigate through surroundings and avoid obstacles while crossing roads[3]. The electronic mobility aids developed for the visually impaired use the ultrasonic waves to detect the obstacles and alerts the user[3]. With intensive research and advancement in technology, we can make use of computer vision which uses artificial intelligence for gaining knowledge from images and imparting visual understanding to the computers[4]. Computer vision can be used to solve the problem of eliminating the nature of being dependent for visually challenged people.

This paper presents a way in which we use the most accurate algorithms, taking care of the trade-off between the speed and accuracy and then combine them under a common platform using raspberry pi so that we have a device that can perform multiple tasks under one roof. Taking into account, the things that truly make a visually challenged individual independent is observed by the fact that he/she should be able to move freely in his/her environment without and assistance i.e. he/she should be able to discover the obstacles in the path and avoid collision with them. Furthermore, there are many signboards on the road and other important things to be read in the vicinity like whether a pedestrian path is there or not or the signal is 'go' or 'stop' or maybe what a shop is selling. Moreover, looking at the other basic features which are very vital to humans is the recognition of currency. To be able to buy and sell goods, the exchange of currency is a must and on top of that, a more crucial factor is to be able to differentiate between real and fake currency.

This paper considers all the above mentioned basic requirements of humans and will include details on how the smart spectacles, which is wearable technology, put on the eyes similar to our normal spectacles, will include three primaries and essential task performed by the device 1) Real-time object detection using the SSD-MobileNet object detection algorithm. 2) Text detection using OCR (Optical Character Recognition) and 3) Currency detection using trained dataset. So the Visually challenged human can get all these benefits in one product itself. This paper also includes how we integrate the different hardware components used in the Smart spectacle including Raspberry pi, a USB camera

Revised Manuscript Received on May 20, 2020.
∗ Correspondence Author

Vraj Baxi∗, Department of Computer, G H Patel College of Engineering and Technology, Anand, India. E-mail: vrajbaxi.1998@gmail.com

Justin Patel, Department of Computer, G H Patel College of Engineering and Technology, Anand, India. E-mail: pateljustin428@gmail.com

Vaibhavi Lakhani, Department of Computer, G H Patel College of Engineering and Technology, Anand, India. E-mail: vaibhavi.lakhani1007@gmail.com

Priyang Bhatt, Department of Computer, G H Patel College of Engineering and Technology, Anand, India. E-mail: priyangbhatt@gcet.ac.in

Module, Headphone with a microphone and SD (Secure Digital) Card.

The next section i.e. Section II focuses on Related Work on the similar technologies which have already been implemented along with their limitations. Section III will describe the proposed framework for the smart spectacles and the implementation of the features it contains. Section IV discusses the results and finally, we conclude this research paper along with the possibilities of future work in Section V.

II. RELATED WORK

A large number of Head-mounted devices (HMD) have been invented for the assistance of the blind and they are very popular wearable devices because people gather information from the surroundings through the motion of the human head[2]. Some of them can allow the visually challenged to read and write the text, some allow them to detect the obstacles, while on the other hand, other good products allow the user to multitask.

Various products already exist in the market but are restricted to their wide acceptance either due to limited features or the high price which cannot be afforded by the small and middle class visually challenged people[5]. For instance, Audio Books give the functionality of reading only. It is too costly as most of them are offered at a subscription price of $ 25 per month and it is not available for all the books. Also, conversion into the audiobook is a tedious task and sometimes handwritten documents cannot be identified properly which is again its limitation[5, 6]. Talking about the screen reader, they facilitate reading Digital format only. Also, they are very costly devices and range from $150 to $1000[5, 6]. The Braille Readers are being used from a long time. They facilitate reading and writing but only tactile materials i.e. materials that can be felt are supported. These devices with limited features are too costly to buy as they range from $1000-$3000[5, 7]. Another interesting device was developed called OrCam MyReader which facilitates reading task only and OrCam MyEye which performs multiple features but this technology is overpriced as they cost around $ 2500 for OrCam MyReader and around $3500 for OrCam MyEye [8]. Esight allows visually impaired people to be able to multitask by re-display of live scenes[5, 9]. Esight is not widely accepted by the price which is around $15000 as well as the fact that it can only be used by people with low vision and not by the people who are completely blind[5, 9]. Some devices also include detecting objects using ultrasonic waves which get reflected on colliding with the obstacles, but with the advancement in learning, we can use deep neural networks to obtain more accurate and efficient results.

The proposed design of the smart spectacles in this paper facilitates multiple features all combined under a single platform including real-time obstacle avoidance by object detection, text detection in the vicinity and currency detection for all types of visually challenged people. The main advantage of such smart spectacles will be the satisfaction of the basic necessities of a human at a very affordable price of $100-$120.

III. PROPOSED FRAMEWORK

Overview

This paper describes the framework and the working of smart spectacles in mainly three parts. The first part focuses on the object detection algorithm, the second part describes how the text will be detected using OCR (Optical Character Recognition), and the third part focuses on the detection of currency.

The main component of the spectacle is Raspberry pi 3, which is a Linux based ARM processor that accepts a micro SD card and allows us to perform multiple functions[5]. For assembling the hardware, we have to put the USB camera in the middle of the spectacle and connect its USB cable to one of the USB slots of the raspberry pi 3 modules. Raspberry pi 3 will get power through the power bank which is connected with it using the power input port of raspberry pi. Our framework involves giving commands to the spectacles through the microphone attached to the headphone and receiving commands from the spectacles through the headphone. The headphone with a microphone is connected to the raspberry pi module with the sound card which has both audio input as well as audio output, attached using other USB port of the raspberry pi module. To use the Google API for Speech to Text conversion, we have to provide internet access which can be done with the hotspot from a mobile device. Fig.1 gives a visual view of how the proposed system will look after the connection of various hardware devices.

Fig. 1 Overall Look

In the proposed design of the smart spectacles, for the convenience of the user, we do not include any switches. To perform any function the user gives commands in the form of Speech and this will get converted to Text by Speech-to-Text Module. While receiving the output the sentence in the form of text gets converted into speech by Text-to-Speech module. Below is the description of Speech-to-Text and Text-to-Speech module.

1) Speech-to-Text Module

Fig. 2 depicts the complete flow in which the input commands which is in the form of speech gets converted into the text form and given as input to perform any task.
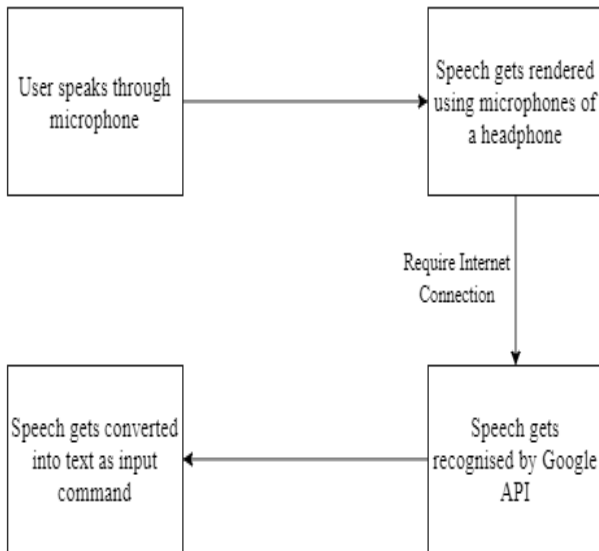
Fig. 2 Speech-to-Text Module

2) Text-to-Speech Module

Fig. 3 depicts the complete flow in which the output obtained by performing any function which is in the form of text is converted back into speech and received by the user. The Python library "pyttsx" works offline and is used to convert Text to Speech[10].
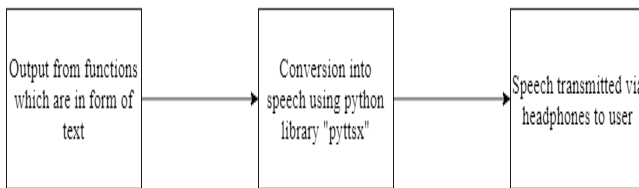

Fig. 3 Text-to-Speech Module

A. Object Detection

There are many algorithms for detecting the objects in the surrounding environment. The traditional algorithm used feature extraction and template matching but we do not achieve high accuracy with such algorithms as they encode low-level characteristics of objects[4]. Using Deep learning object detection algorithm, we have Faster R-CNN's, SSD (Single Shot Detectors), YOLO (You Only Look Once)[11]. MobileNets[12] network architecture can be used inside the object detection pipeline but these heavy architectures are unsuitable for devices like Raspberry Pi[4]. Hence we combine MobileNet architecture and SSD framework to arrive at a very fast and efficient real-time object detection algorithm[4]. The SSD predicts different scale feature maps to gain high accuracy on predictions made by it and this also improves its capabilities of detecting different size objects[13]. Fig.4 determines the architecture of SSD and YOLO. Also, the faster working of SSD as compared to the Faster R-CNN model was due to the subsampling of the image[13]. It has seen that SSD works faster and had superior performance than YOLO[13].
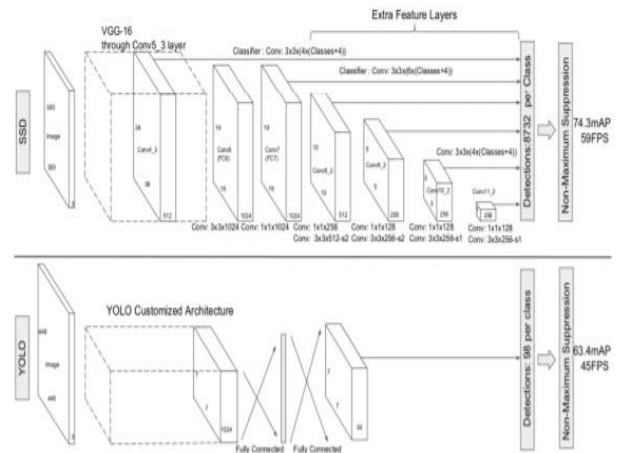

Fig.4 Comparison between SSD and YOLO Single Shot Detection object detection[14].

The performance is better in SSD as it uses varying size convolutional networks in comparison to the two connected layers used by YOLO[13]. The model loss obtained during the training of the model is weighted sum of the localisation loss (box) and confidence loss (label) as described in the equation 1 below[13, 14].

$$L(x,c,l,g) = 1/N(L_{conf}(x,c) + \alpha L_{loc}(x,l,g)) \qquad (1)$$

Where, N is the number of matched default boxes, the localisation loss is a Smooth L1 loss between the ground truth (g) box and the prediction box (l)[13, 14]. For determining some close objects out of all the objects detected by SSD-MobileNets object detection, we can use the Triangle Similarity method, but, we require the knowledge about the width of every different object in which we're interested and this is not possible while dynamically determining objects. Hence, in the module that we develop we can detect certain smaller or larger objects through utilizing visualize boxes and labels on the image array. In order to calculate the width of the detected object, we measure how many pixels-wide the object is. We can do this with subtraction of box indexes. Then we'll round the obtained number and we get the approximate distance. If the distance is less than 5 meters then we are going to warn the visually impaired person through speech. Fig. 5 depicts the complete flow of how the smart spectacles detect the objects very quickly and accurately using SSD-MobileNets architecture and gives the output to the user.
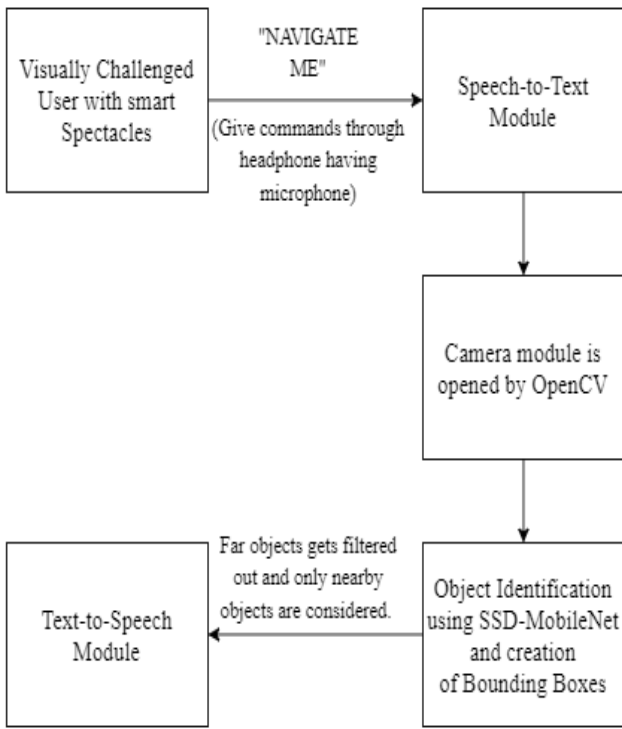
Fig.5 Object Detection Module

This flow chart can be modified if the user wants to find the existence of specific objects in the vicinity. In this scenario, the algorithm compares the result of all the labels associated to various bounding boxes with the desired object to be found and return YES (if the object is found) or NO (if the object is not found) through speech. Fig. 6 given below depicts the complete process.
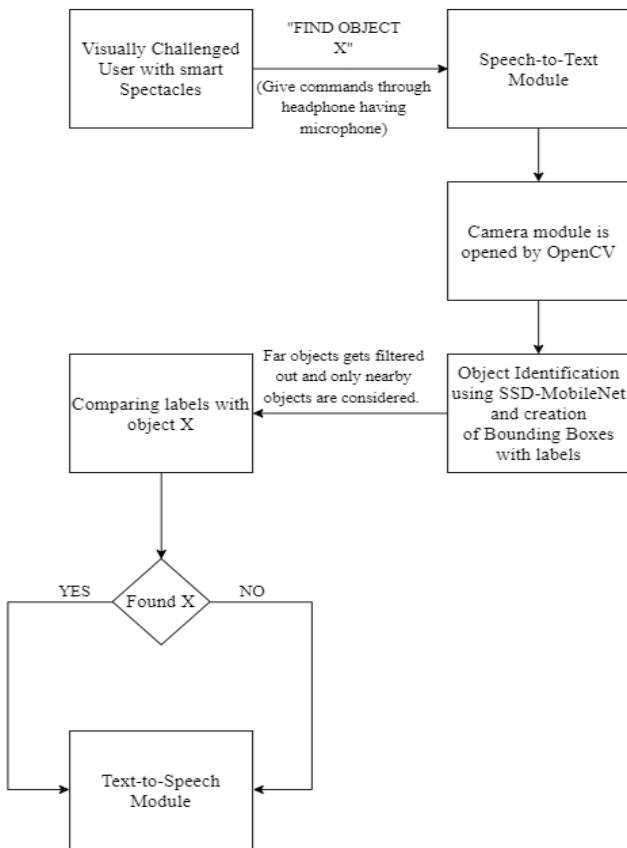

Fig.6 Specific Object Detection Module

B. Text Detection

The second vital function performed by the smart spectacle is detecting the Text in the vicinity. Detection of texts is done by Optical Character Recognition (OCR), which is the technique used to convert printable text to editable text[15]. OCR allows a machine to detect the text automatically[15]. This method is highly dependent on several factors such as image size, complex backgrounds, style, orientation, etc. We have used software for OCR named Tesseract[15]. It is open-source and platform-independent software written in C++ Language making it highly portable and platform-independent[15]. Tesseract works in five folds which has been described in the following lines. The first step is the conversion of an image into binary images using the Adaptive Threshold method[16]. The second step is connected component analysis[15] for extracting character outlines. This method is very significant as it does the OCR of the image with a black background and white text. Tesseract Software is the first most software to provide this kind of processing. Furthermore, obtained outlines are converted into Blobs[15]. The Blobs are organized into text lines and regions which are analyzed for some fixed area or equivalent text size. This text is divided into words using definite spaces and fuzzy spaces[15]. In the next step, the process is made into a two-pass procedure. In the first pass, recognition of every word takes place and passed to an adaptive classifier as training data. In the subsequent pass, various issues during the extraction process are resolved. In the case of a complex color image, for obtaining the accuracy as compared with a grayscale image (without color), there should be pre-conversion of a color image into a grayscale image and then OCR Technique should be applied. Fig. 7 depicts the complete flow.
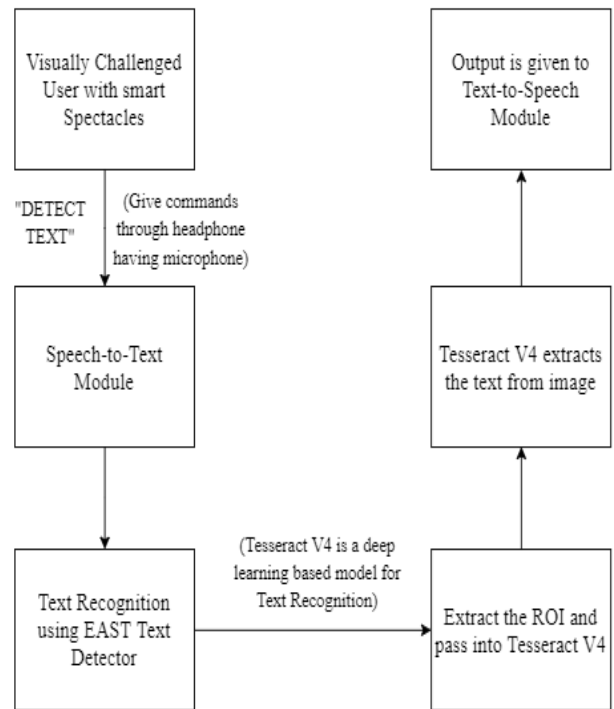

Fig.7 Text Detection using OCR and Tesseract V4 Software

C. Currency Detection

This feature recognizes the currencies from the trained dataset using MobileNet. Using the above mentioned speech-to-text module, command from the user is converted into text which is then fed as input in this currency detection module. By analyzing the training dataset, more accuracy can be obtained and improved according to our expectations. In this feature, there would be an adequate number of images in the dataset which can be used for training purpose. Initially, multiple images are captured until the appropriate image is found for further processing. Hence it follows two subtasks: positioning and classification[17]. This is followed by the feature extraction feature which describes location regression and object classification[17]. This allows determining the match-proportion. Match-proportion gives the value which shows the statistical data matched interesting key points. These key points are selected by analyzing edges, corners, or blobs and are compared with the trained model. The output is obtained which has the highest value of match-proportion. Fig 8 shows the above-mentioned flow of the process.
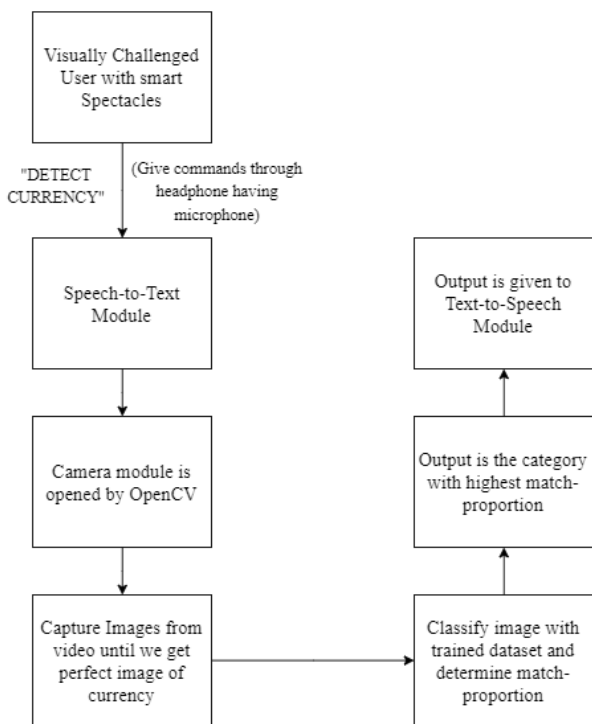

Fig.8 Currency Detection Module

Table 1 given below compares various devices available with the proposed design of the spectacle.

Table 1 Comparison of various devices and the proposed design of the Smart Spectacle

Device	Function	Remarks	Price
OrCam MyReader[8]	Reading Text	Only feature of reading texts	$2500
OrCam MyEye[8]	Using Artificial Intelligence it reads text, recognizes faces, and identifies products.	Very costly which is mostly unaffordable by all low and middle class visually	$3500
		challenged people	
White Cane (Identification Cane and Support Cane)[18]	Alerts people in the surroundings through the design of the cane.	Does not provide identification of obstacles or currency or text and user needs to change the grip.	$18
White Cane (long Cane)[18]	Alert user of terrain and height changes and obstacles	Not always accurate and not very secure. Does not provide identification of obstacles or currency or text in the surroundings.	$75
Smart Cane[18]	Emits ultrasonic waves to detect the obstacles in the surroundings and alerts the user by vibrations. Uses arduino programming	Does not provide identification of obstacles or currency or text in the surroundings. Hence provides very limited features. And user needs to carry everywhere they want to go.	$70
Sonicguide – a binaural sonic guide[19]	A pair of spectacle frames with one ultrasonic transmitter on one side and a receiver on other side. The resulting interaural difference determined the obstacles.	Does not provide active scanning of the environment. Reduced speed and very costly.	$3000

Smart Spectacles (Proposed Design)	Offers Multiple features including Real time obstacle avoidance by detecting object in the surrounding, Detects text in the vicinity and reads it through Headphone and Detects Currency for the visually challenged.	All the basic necessities of a human are satisfied by the help of this smart spectacle which can be made at a very affordable price.	$100-$ 120

We find that the smart spectacles are the best in comparison with all the available devices as it provides multiple features at a very affordable cost without compromising efficiency.

IV. RESULTS

We have tested the various modules described in this research paper. We found all the modules to be very efficient and quick in producing the results.

For testing the object detection module using SSD-MobileNets, we used a bottle. The object detection module listed in the paper accurately classified the bottle with 100% confidence. Fig.9 shows the result of the label and the confidence generated before it gets converted to audio output.
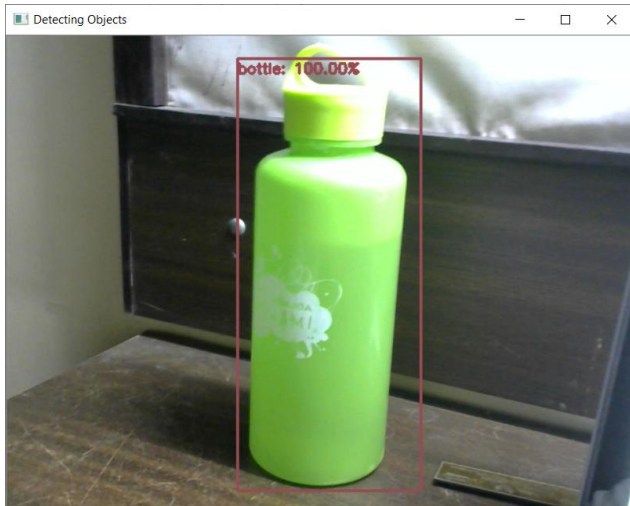

Fig.9 Result of detecting a bottle using SSD-MobileNets

The intermediate result of the Text detection module before getting converted into audio output is shown in Fig. 10.
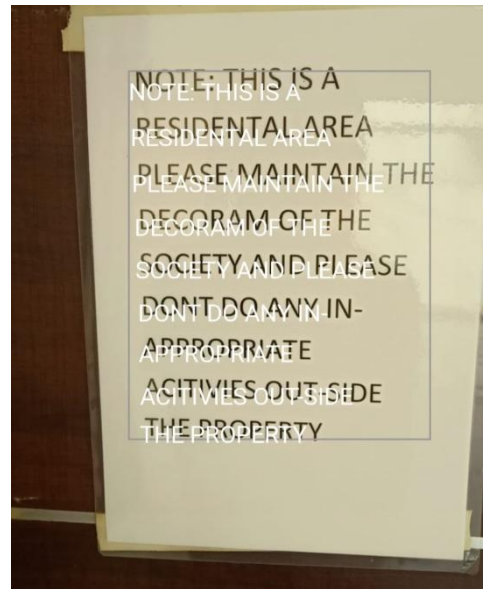

Fig.10 Result of Text Detection Module

The currency module used in the smart spectacle also portrayed correct output. We used the Indian currency denominations of Rs. 20 and Rs 50 and received correct classification of the currency as shown in Fig.11.
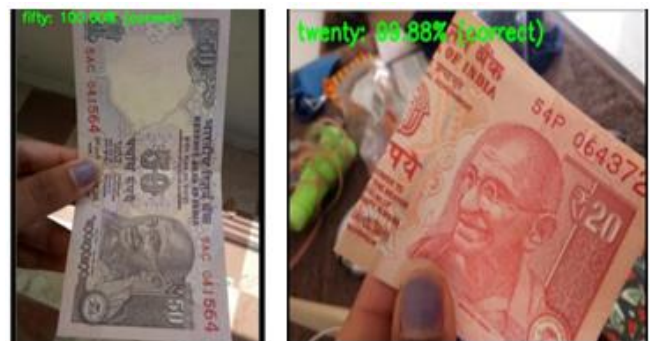

Fig.11 Results of Currency Detection Module

V. CONCLUSION AND FUTURE WORK

Through the writing of this paper, we have tried to help the visually challenged people by proposing a universal solution which satisfies all their basic needs of moving freely in the surrounding by detecting the obstacles in their vicinity, detecting what is written on the roads and shops along with being able to read different signboards and finally detecting currency so that no individual can cheat them while trading for goods. Although some of the features might be available in other products available in market, but all the described features are not available in a combined form on a single platform and also the available devices which facilitates multiple tasks are were very costly for the visually challenged people to purchase as they usually belong to low and middle-class families.

In the future, much more work can be done to improve the efficiency of the algorithms which helps in faster and more efficient detection of the objects in real-time. Also, we could work upon making the Text-to-Speech Module work completely offline and give the same accuracy as it gives online. More features can be added to this smart spectacle like the emergency tracking

button which sends the location of the visually challenged person to his family person so, if he/she is having some problem then through GPS the location is tracked and send to the close family members or else, if he/she is in danger then the location can be sent directly to the police. Another, exciting feature that can be included in the sensors on the spectacles that vibrate and produce sound when there is a substantial change in the position of the spectacle so that, if by chance the spectacle falls or misplaced by the visually challenged they can locate it by hearing the sound.

REFERENCES

1. Blindness and Visual Impairment: Global Facts: https://www.iapb.org/vision-2020/who-facts/
2. R. Velázquez, "Wearable Assistive Devices for the Blind". Chapter 17 in A. Lay-Ekuakille & S.C. Mukhopadhyay (Eds.), "Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization", *LNEE 75, Springer*, pp 331-349, 2010.
3. Illinois Library: Blind/Visual Impairment: Common Assistive Technology: https://guides.library.illinois.edu/c.php?g=526852&p=3602299
4. Sejal Gianani, Abhishek Mehta, Twinkle Motwani, Rohan Shende. "JUVO - An Aid for the Visually Impaired", *International Conference on Smart City and Emerging Technology (ICSCET)*, January 2018.
5. Esra Ali Hassan, Tong Boon Tang. "Smart Glasses for the Visually Impaired People", from book Computers Helping People with Special Needs: *15th International Conference, ICCHP* 2016, Linz, Austria, July 13-15, 2016, Proceedings, Part II (pp.579-582).
6. The Macular Degeneration Foundation, "Low Vision Aids & Technology", Sydney, Australia: The Macular Degeneration Foundation, July 2012.
7. Velázquez, R. "Wearable assistive devices for the blind. In: Lay-Ekuakille, A., Mukhopadhyay", S.C. (eds.) Wearable and Autonomous Systems. *LNEE*, vol. 75, pp. 331–349. Springer, Heidelberg (2010).
8. OrCam, OrCam. http://www.orcam.com.
9. Esight. http://esighteyewear.com/
10. Python Package Index: https://pypi.org/project/pyttsx3/2.7/
11. Yundong Zhang, Haomin Peng, Pan Hu. "Towards Real-time Detection and Camera Triggering". Available: http://cs231n.stanford.edu/reports/2017/pdfs/808.pdf
12. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". Available: https://arxiv.org/pdf/1704.04861.pdf
13. OSCAR ALSING, "Mobile Object Detection using TensorFlow Lite and Transfer Learning", STOCKHOLM, SWEDEN 2018, https://kth.diva-portal.org/smash/get/diva2:1242627/FULLTEXT01.pdf
14. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, and Scott E. Reed. "SSD: Single Shot MultiBox Detector." In*: CoRR* abs/1512.02325 (2015).
15. SMITH, R.,"An Overview of the Tesseract OCR Engine. In proceedings of Document analysis and Recognition.", *ICDAR 2007. IEEE Ninth International Conference*.
16. Y. WEN, Y. L. 2011., "An Algorithm for License Plate Recognition Applied to Intelligent Transportation System.", *IEEE Transactions on Intelligent Systems*, pp.1-16.
17. Zhang, Q., & Yan, W. Q. (2018), "Currency Detection and Recognition Based on Deep Learning", 2018 *15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*.
18. Whitney Huang, Hunter McNamara, Diana Molodan, Amol Pasarkar,"Smart Cane"," https://soe.rutgers.edu/sites/default/files/imce/pdfs/gset-2014/Smart+Cane+Final.pdf".
19. L. Kay, "A sonar aid to enhance spatial perception of the blind: Engineering design and evaluation", *Radio Electron. Eng.,* vol. 44, no. 11, pp. 605-627, 1974.

AUTHORS PROFILE

Vraj Baxi is pursuing Bachelor of Engineering in Computer Engineering from G.H. Patel College of Engineering and Technology (GCET) located in Vallabh Vidyanagar, Gujarat. He is currently studying in Semester 8 of his curriculum. Vraj has been the member of various professional Committee including IEEE, Indian Society of Technical Education and Computer Society of India. He has been an active volunteer and coordinated various events at state, national and international level. He was the secretary of Computer Society of IEEE GCET Student Branch. He has won many awards during his undergraduate study including SHRI DEWANG MEHTA IT AWARDS 2019, Second position in technical event at national level Technical Festival IMAZE'19, First position in Inter college level presentation competition and many more. He has been the Class Topper in his 4th and 5th Semester and the Institute Topper in his 7th semester during his undergraduate study. He has been working on various projects of varying objectives including "Portal for Social Enthusiast", which aimed at connecting socially enthusiastic volunteers, "University Leave Portal", which aimed at helping university professors to easily take leave by making the complete system online, "Portal for international Conference on Recent Advances in Computer Science and Technology" and "Blind's Spectacles". Vraj believes that it is very important to maintain balance between academics and extra-curricular part of life.

Justin Patel, currently studying in 8th semester B.E. in Computer Engineering in G.H. Patel College of Engineering and Technology (GCET) located in Vallabh Vidyanagar, Gujarat. He has been the member of Computer Society of India (CSI) for one year. He has participated in several national and international events, conference, workshops and hackathon. He has an android mobile application on Google Play Store with 10,000+ downloads and 4.6 rating. He has also won many coding competitions and been a coordinator of Android committee at national level tech-fest. He has worked upon several projects such as "Gadget Detox", "University Leave Portal", "Blind's Spectacle" etc. His area of interest is critical, analytical and logical programming, Machine Learning and Data Science.

Vaibhavi Lakhani, currently pursuing a Bachelor of Engineering in Computer Engineering from G H Patel College of Engineering and Technology (GCET) in Vallabh Vidyanagar, Gujarat. She has been an active member of various professional societies such as the Institute of Electrical and Electronics Engineers (IEEE) and Computer Society of India (CSI). She has diligently participated, volunteered as well as coordinated many events and competitions at state, national and international levels. Furthermore, she has also proved her competence by participating and winning in several technical competition in the field of programming. Moreover, she has contributed to many projects in her undergraduate study including "Time Management Application", "University Leave Portal" and "Android Music Application". Besides, her area of interest includes Big data, Databases and knowledge-bases, Machine Learning and Data Science.

Priyang Bhatt (Assistant Professor) joined GCET in 2007. He pursued his B.E in Computer Engineering from Charotar Institute of Technology (Now CHARUSAT), Changa, Gujarat University and M.E. in Computer Engineering from Birla Vishwakarma Mahavidyalaya, Sardar Patel University. He is currently pursuing his Ph.D. from Gujarat Technological University in the area of Securing Internet of Things (IoT).

He is CCNA, RHCE and RHCSA certified. He has developed and published an Android app for first-year students with the name "CPU EXAM PREP" on the Google Play Store. he has conducted the number of workshops and delivered expert talks on the area of Machine Learning, Python Programming, Android Programming, IoT and CISCO Networking. He has published several papers in International Journals and Conferences. He is involved in teaching subjects like the Internet of Things (IoT), Android Programming, Parallel Processing, Wireless Communication and Network Security of Computer Engineering/IT at UG and PG Level.