
Assisting visually impaired people by Computer Vision- A Smart Eye

Arun Kumar Ravula^{1*}, Palapati Vasavi², K.Ram Mohan Rao³

^{1*}Department of Computer Science and Engineering, Osmania University College of Engineering, Osmania University, Hyderabad, Telangana, India,

²Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India ³Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, Telangana, India,

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract:

According to statistics of WHO (World Health Organization), there are nearly 1 billion visually-impaired people worldwide and among these majority of them are completely blind. The proposed system will help such people to guide through their path, which does object detection and object recognition using image processing and deep learning techniques. It also provides distance of that object from the camera. All the detected objects are speech processed using text to speech API and are speak out to the user for assistance. These 3 features can be provided through a single phone, which reduces costs and complexity, and enhances the practicality of the application. The results obtained from the proposed application boosts the confidence level and comfort level of the blind; in this way they become more independent.

Keywords: Deep learning, Object Detection and Recognition, Language processing, Computer Vision, Object Distance Estimation

1. Introduction

The numbers of visually challenged people are growing rapidly from past few decades. According to World Health Organization, about 1031[1] million individuals are known to be visually challenged globally. The proposed system will help such people to guide through their path, which does object detection and object recognition using image processing and deep learning techniques. Assistive, wearable devices[10] are most useful because of minimum use of hands. The assisted vision device will be worn by the person, and there will be a camera in it, which detects things that come in front of him. Using the device is easy, it detects the things and tells the user what is the object which is present in front of him/her through a speaker and it even tells how far the object is from the person viewing or the user who is using it.

The proposed work is simple and user friendly, assisted vision system is designed and implemented in order to improve the movement of both blind and visually challenged people in any area, this system helps them to move easily by saying the distance of the object from user, which is estimated by the system. And the person can be able to avoid accidents as the system will warn him if any obstacle is approaching towards him.

The main aim is to help visually impaired people, by providing them better services with less cost; So that all of them can easily afford and also through this device provide them safety during they walk on roads or any other places when they are walking. In this device, we introduced a system in which a camera is placed on blind person's glasses, the task of the camera is to take images every second and process them and give output. It does object detection to say the user what is the object present in front him making it easy to identify the objects for visually challenged people. Passing from an unknown area or environment is a real challenge for visually challenged people, when they can't depend on their eyes. Since objects may be moving, the visually challenged people will hear the noise which is usually produced while moving, blind people develop the ability of listening and hearing to locate the objects. However they can't depend more on their senses of touch and hearing when the matter is to determine where the object is located exactly. The device which we developed will do it for them without touching the objects. The system will process the image and identify the objects in it and speak them out to the user including the distance of the object from the user.

2. Related work

There are many systems which are developed to help the visually aided people and improve their quality of life. Many such devices or systems have certain limitations. People with moderate or severe vision loss need better positioning and control of their personal belongings, so a feasible solution is proposed to help them handle the objects used daily.

In the paper [7] J. Ying et al. proposed a Navigation Device for assisting visually impaired people. They proposed a new approach, namely, DLSNF [1] based on the YOLO Architecture. A real world dataset with 4700 images is used for DLSNF .

Megha P Arakeri et al.[2] proposed system captures the visible information towards the user, identifies the text in the image and speaks it out. It shows the object distance from the user and speaks the details about object. The product majorly helps the user to understand the object details, any readable information which surroundings him. The proposed system is a user-friendly application, portable and compact. The main drawback of this system is it is designed to address the raspberry pi, ultrasonic sensor, power resource, NoIR camera , breadboard, resistors, jumper wires and the press buttons.

L Dunai et al[3] suggested a system called ‘CASBlip’, a wearable device, Cognitive Aid System for Blind People. The objectives of the device are an obstacle detection , orientation and navigation.

Roentgen, Uta & Gelderblom [4] made review on “Inventory of Electronic Mobility Aids for Persons with Visual Impairments”. In this review they classified 21 available with required functionalities out of 146 a electronic mobility aids,12 for object detection and 9 were aimed at navigation . They summarized recent developments in a database with 17 attributes. They Delivered as practical information on various electronic mobility aids, these can help the adequate knowledge for both users and manufacturers. They suggested users to use the personalized assisted devices depending on their individual requirements.

Zhiheng Yang et al. [5] proposed a Detection system for pedestrian for autonomous driving vehicles. The system architecture is similar to Darknet, is one of the state-of-the-art Deep Learning algorithm for object detection. The proposed model is able to predict the location of the object by the anchor boxes convolution. They have used several strategies for improving the accuracy like Multi-scale training, hard negative mining, Model pre-training and usage appropriate key parameters.

J. M Loomis., et al. [6] proposed a system “Makino” an ETA system. The system communicates user and computer database by using a digital mobile phone. User position is transmitted by GPS through mobile phone via designated voice.

Zhiheng Yang.,et al. [8] developed a system for object detection called” Real–Time Assistance Prototype”. The proposed system is able to detect the objects in real time static and movable based on stereo-vision technology. The system is also responsible for offering the 3-D information related to the object environment.

Imene OUALI, et al.[9] presented a new architecture to find the drug information and remaining number of pills for visually impaired people. They designed AR architecture for object recognition for visual impaired people needs.

Milios Awad, et al.[11] proposed an android mobile application “Intelligent Eye” aims visually impaired people and designed with several features like Detection of light, Color, Banknote and Object etc. In future they may add features such as Barcode, Personal Card Scanner and traffic light Detector.

3. Methodologies

The contributions of this proposed work are 1. Object detection, 2. Obstacle distance estimation, 3. Speech processing. Figure 1 represents the architecture of Smart Eye application.

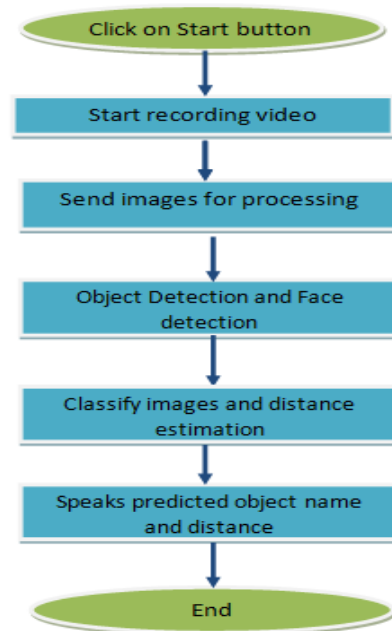


Figure 1: Assisted Vision activity diagram

3.1 Object Detection

The most important module of the device is object detection. It is a technique for locating the objects. There are several algorithms in deep learning and machine learning to do the object detection. Here we will train the device to detect objects, thus when a particular object comes in front, it will recognize the object. Thus detecting objects will make it easier to understand what type of object is in front of user

Faster-RCNN

Faster R-CNN uses several numbers of innovations to increase training and testing speed while also improving detection accuracy. It is faster than R-CNN and reaches higher mAP on PASCAL VOC 2012. It is also faster than SPPNet.

SSD

Single shot Detection [15] is uses a single feed forward convolutional network architecture to detect classes without requiring second level classification. It removes bounding box proposals like which are used in R-CNNs, includes a progressively reducing convolutional filter for detecting object classes and offsets in bounding box locations. It has a base network VGG-16 followed by multi box convolutional layers. SSD uses 8732 bounding boxes for best coverage of scale, aspect ratio and location. It never considers the object which is containing the confidence score lower than 0.001. It uses the loss function, which contains confidence loss and localization loss.

The actions performed for the object detection are

1. For input, we need to take an image to detect objects in it.
2. The image is divided to form different regions.
3. Then each region is considered as separate image.
4. All the images(regions) are given to CNN[17] in order to classify each one of them into various

classes.

5. After dividing the image into their corresponding classes, the original image can be rebuilt from the regions which have the detected objects.

3.2 Distance Estimation

Detecting objects and saying what is that object is not enough for a visually challenged person, because what is the use of saying what is that object without saying at how much distance is it there from the user. So distance calculation is more vital. Example: Detecting a car is not important, but saying where the car is from the user, how much distance is it from him is important. If the car is very near user can take precautions and move aside. Thus distance estimation along with object detection is most important.

In order to obtain the distance of object from camera we are using the size of the object to find it, if an object approached closer to cam, then its size will increase, by using this technique we are calculating the distance of object using certain parameters". The exact distance value is not important for the user, so we generalize them to certain range like "very near" or "near".

3.3 Speech Processing

Detecting the objects is first step, calculating the distance of the object is second step, the third step is to say user what the object is and where the object is either far, near or very near. User can understand what the object is and can react accordingly. This will make our device most effective

The flow of execution is followed as below steps:

1. Importing all the required libraries.
2. Selecting the required model.
3. Load the model which we downloaded is tensor flow model
4. Start the video capture from the camera.
5. Run the tensorflow session
6. Object distance will be calculated, the exact distance value is not important for the user, so we generalize them to certain range like "very near" or "near"
7. Finally, the output should be spoken out so that user can listen and understand it.

4. Results and Discussions

This section presents the results achieved by Smart Eye application. The experiment results were achieved through trained dataset from Microsoft Coco:[12],[13] ,used Fast-RCNN[14],[20]. Figure 2 shows the starting button of proposed application. Every object that has been detected will have a score assigned to it, this score will show how accurate the given object when matched with the trained object. Here we are not considering objects have score less than 70% (0.7). This will help us to get more accurate results. In the figure 3 the camera is detecting the object i.e bed and chair placed in-front of it. We can also see the score of the object which is the accuracy of the object detection. For the object bed the score is 70% and it is greater than the object chair score i.e 69%. Figure 3, 4, 5 and 7 shows the objects identified by the Smart Eye application i.e Scissors, cup, remote and Phone. Proposed application also able to detect the face, In figure 6 a person is detected with 98% accuracy. For the better results Mobilenets[16] are preferable.

Smart Eye application is also responsible for announcing detected objects with clear voice. It also announces "the object is very nearer" or "the object is very far". Here we are taking maximum apex distance of 0.5; if apex distance is less than 0.5 then it means that the object is approaching very close to us.

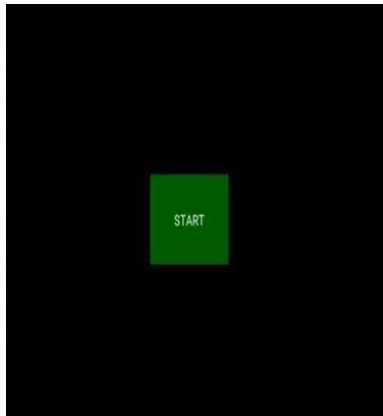


Figure 2: Start Button



Figure 3: Detecting objects



Figure 4: Detecting Scissors, cup



Figure 5: Remote



Figure 6,7: Detecting Face with 98% and Cell Phone with 75% accuracy

5. Conclusion

Distance estimation was the most important feature as it is most useful for the user. When user uses the device, the object which is placed in front of him will be detected and spoken out to the user along with the distance of the object from him. This will make it more useful because, now he can know where the object is present in-front of him rather than just knowing what object is present in-front of him.

“Assisted vision project helps visually challenged people to identify objects, persons who are in their visible range. It uses object detection and object recognition, distance measurement and can be able to speak out the object that is being detected in speech that helps them to identify those objects and make their way by listening to it. As reported by the world health organization the stats state that about 285 million individuals worldwide are estimated to be visually challenged. This is the reason why these types of applications are of really great help”.

“We expect to implement this system into glasses used by visually challenged people.

As of now the system is being showcased using mobile phone. In future we will develop this into glasses. We also expect to make it useful in saying directions to the user. Our system will ask user to move left or right direction depending up on the object coming towards him. Example: if a car is approaching very close towards the user, the system will say user to move away in specific direction making it easier to escape from those kinds of situations”.

References

- [1] World Health Organization (WHO). 2014. Visual Impairment and Blindness. <https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [2]. M. P. Arakeri, N. S. Keerthana, M. Madhura, A. Sankar and T. Munnavar, "Assistive Technology for the Visually Impaired Using Computer Vision," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1725-1730, doi: 10.1109/ICACCI.2018.8554625.
- [3] L. Dunai, B. D. Garcia, I. Lengua and G. Peris-Fajarnés, "3D CMOS sensor based acoustic object detection and navigation system for blind people," IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, Montreal, QC, 2012, pp. 4208-4215, doi: 10.1109/IECON.2012.6389214.
- [4] Roentgen, Uta & Gelderblom, Gert Jan & Soede, Mathijs & Witte, Luc. (2008). Inventory of Electronic Mobility Aids for Persons with Visual Impairments: A Literature Review. *Journal of visual impairment & blindness*. 102. 702-723. 10.1177/0145482X0810201105.
- [5] Z. Yang, J. Li and H. Li, "Real-Time Pedestrian Detection for Autonomous Driving," 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), Singapore, 2018, pp. 9-13, doi: 10.1109/ICoIAS.2018.8494031.
- [6] J. M Loomis., R. G. Golledge, & R. L. Klatzky (2001). GPS-Based Navigation Systems for the Visually Impaired. Pp. 429-446.
- [7]. J. Ying, C. Li, G. Wu, J. Li, W. Chen and D. Yang, "A Deep Learning Approach to Sensory Navigation Device for Blind Guidance," 2018 IEEE 20th International Conference on High Performance Computing and Communications; United Kingdom, 2018, pp. 1195-1200, doi: 10.1109/HPCC/SmartCity/DSS.2018.00201.
- [8] L. Dunai, G. P. Fajarnes, V. S. Praderas, B. D. Garcia and I. L. Lengua, "Real-time assistance prototype — A new navigation aid for blind people," IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, 2010, pp. 1173-1178, doi: 10.1109/IECON.2010.5675535.
- [9] Imene OUALI, Mohamed Saifeddine HADJ SASSI, Mohamed BEN HALIMA, Ali ALI,"A New Architecture based AR for Detection and Recognition of Objects and Text to Enhance Navigation of Visually Impaired People",*Procedia Computer Science*, Volume 176,2020,Pages 602-611,ISSN 1877-0509,<https://doi.org/10.1016/j.procs.2020.08.062>.
- [10] Jafri, R. et al. "Computer Vision-based Object Recognition for the Visually Impaired Using Visual Tags." (2013).66.
- [11] M. Awad, J. E. Haddad, E. Khneisser, T. Mahmoud, E. Yaacoub and M. Malli, "Intelligent eye: A mobile application for assisting blind people," 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), Jounieh, 2018, pp. 1-6, doi: 10.1109/MENACOMM.2018.8371005.
- [12] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *Computer Vision—ECCV 2014*. Springer International Publishing, 2014. 740–755.

- [13] Russakovsky, Olga; et al. (2015). "Imagenet large scale visual recognition challenge". *International Journal of Computer Vision*. 115 (3): 211–252. arXiv:1409.0575. doi:10.1007/s11263-015-0816-y. hdl:1721.1/104944. S2CID 2930547.
- [14] Girshick, Ross (2015). "Fast R-CNN" (PDF). *Proceedings of the IEEE International Conference on Computer Vision*: 1440–1448. arXiv:1504.08083. Bibcode:2015arXiv150408083G.
- [15] Liu, Wei (October 2016). "SSD: Single shot multibox detector". *Computer Vision – ECCV 2016. European Conference on Computer Vision. Lecture Notes in Computer Science*. 9905. pp. 21–37. arXiv:1512.02325. doi:10.1007/978-3-319-46448-0_2. ISBN 978-3-319-46447-3. S2CID 2141740.
- [16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. *Mobilenets: Efficient convolutional neural networks for mobile vision applications*. arXiv:1704.04861, 2017.
- [17] K. Simonyan, and A. Zisserman. "Very deep convolutional networks for large-scale image recognition." In *ICLR*,
- [18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions." In *CVPR, 2015* G. Xie, and W. Lu, "Image Edge Detection Based on OpenCV." *International Journal of Electronics and Electrical Engineering* 1 (2): 104-6, 2013
- [19] W., Liu, D., Anguelov, D., Erhan, C., Szegedy, S., Reed, C. Y., Fu, & A. C., Berg. "Ssd: Single shot multibox detector." In *European conference on computer vision 2016* .pp. 21-37
- [20] S., Ren, K., He, R., Girshick, & J., Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks." In *Advances in neural information processing systems*.2015. pp. 91-99.
- [21] J., Dai, Y., Li, K., He, & J. Sun, "R-fcn: Object detection via regionbased fully convolutional networks". In *Advances in neural information processing systems 2016*. pp. 379-387.
- [22] R. Girshick, "Fast R-CNN," 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *European Conference on Computer Vision (ECCV)*, 2014.
- [24] J. Huang et al., "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors," 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 3296-3297, doi: 10.1109/CVPR.2017.351.