Al Based Advanced Navigation Assistant for the Visually Impaired

Gayitri H M[1*], Xavier Fernandis[2], Shrikant Malgai[3], Ujwal Naik[4], Shobit Naik[5], Kumar J R[6]

[1,2,3,4,5] Department of Electronics and Communication Engineering, JSS Science and Technology University

(Formerly Sri Jayachamarajendra College of Engineering,) Mysore 570006, India

[6]Department of Bio Chemistry, Faculty of Life Sciences , JSS Academy of Higher Education and Research , Mysore 570015, India

*Corresponding author: gayitrikumar@sjce.ac.in

Article Info
Page Number: 2754-2764
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract

Vision plays an important role in our existence. without the assisted navigation in outdoor environments the visually impaired people face a lot of problems. This paper presents a system that helps the blind people to navigate without much difficulty. The system incorporates a voice-based user interface where the user interacts with the system by employing a microphone and listens to the instructions through speakers. The system also incorporates raspberry pi as a mini-computer. To feed the user with the appropriate directions to reach his destination the Google Maps API is interfaced into the system as well. The system also detects the obstacles and lets the user know the location and the distance of a particular object. The YOLO model is employed to detect the object and the DisNet algorithm model is built on top of the YOLO algorithm to get the distance of the object detected from the YOLO model. The system is multi-programmed to run the modules simultaneously to attain real-time behavior. The user can query the system and ask for the estimated distance from the current position to the desired destination, the estimated time of travel, and the objects present in the frame of reference.

Keywords: -YOLO, DisNet, Raspberry PI, obstacle detection, Google Maps API

1. Introduction

Visual disability is one of the most significant types of disabilities. Recent statisticsfrom the WHO estimate the number of visually impaired to be around 2.2 billion. Many people agree that vision plays an important role in our lives there are numerous challenges when moving from a particular source location to a destination in an outdoor environment. Although, it might be comparatively easy for the visually impaired to navigate in a well-known environment like on their home and home streets because they havebeen living there since birth. However, it is difficult for the visually impaired to navigate into an unfamiliar outdoor environment. The visually impaired wants to know their localization and the objects that are in their surrounding environment and more importantly the objects which are in front of them. A white caneis mostly used by the visually impaired tonavigate around their surroundings,although the use of the canes helps in some way it doesn’t provides any information about the obstacles that are approaching or the obstacles that are not tapped by the impaired.

Recent improvements in technology have proposed different types of devices and electronic travel aids which is a combination of various sensors which is used to map the environment and to provide the voice instructionsthrough speakers but these electronic-based devices don’t give anyreal-time output and as thereare many sensors involved the cost is more. The electronic travel aids provide with the object detection and obstacle avoidance system for safe the movement of the visually impaired. The electronic travel aids works the sensors are integrated tocollect information from the surroundings and then the information is processed using a CPU and detects objects and gives the user voice feedback based on the detections. The ultrasonic based devices can detect an obstacles in the range of 250 cm bygenerating a signal of frequency of 4KHz and it receives the reflected sound signal from the obstacle and the distance is measured based on the time of reception and the sound signal pulse. In some system spectacles and shoes are interfaced with the electronic sensors which aids the visually impaired. The problem with using the ultrasonic devices is that is it doesn'tgive proper reading if the object absorbs the sound waves or if there are corners wherethe sound wave would bounce off the surface. Furthermore, any object information isn't derived from the ultrasonic module.

Due to the advancement of computing capability available storage and CPU frequency are powerful enough to solve the real world problems.Hence we like to develop a system which assists the visually impaired in unknown surroundings. While the visually impaired is navigating to a particular destination if location of the visually impaired is known, then the user has to know the instructions to reach to local waypoints on the way to his destination, the visually impaired wants know about obstacles and other persons if they are present around.The Object recognition module solves this problem by detecting the object and it's location through bounding. We have built a system which is robust, cost effective and easy to use blind navigation system that consists of a camera,raspberrypi,GPSmodule,microphone and a speaker. Which guides the visually impaired to navigate to their destination.

The proposed system enables a visually impaired person to travel unassisted. This system integrates various electronic sensors and complex algorithm to build a robust wearable navigation system. The proposed system detects dynamic real-time objects in real-time and provide voice instructions to the visually impaired.

2. Methodology

The basic aim of the project is to provide a sense of safety and comfort to the visually impaired to navigate their surroundings easily without taking the help of others. The project intends to make use of state of the art deep learning concepts to help the user navigate from one point to another and also indoors. It also aims to detect and tell them the position of the object they want to interact with. The basic structure of our project is shown in figure 1.

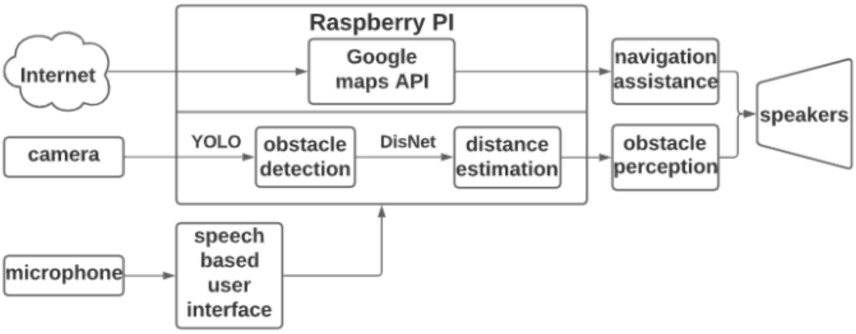

Figure 1: proposed block diagram

The user interacts with the system using the microphone the user can give appropriate commands to start,stop or change the state of the system when the user start the obstacle avoidance system the camera sensor starts capturing the surrounding and passes the frame to the YOLO object classifier the classifier detects the obstacles and passes the output to the DisNet algorithm which measures the distance between the user and the object based on its output appropriate instructions are given to the user to avoid the obstacles.

The User can start the navigation system by interacting with the microphone when the user sets his destination the current location of the user is extracted using the GPS module and the request for the waypoints is sent to the Google maps API this API responds with a Json with the latitude and longitude of the waypoints and the necessary instructions along with it.The navigation system stores this response as a queue the user gets the appropriate instruction to reach for his destination.The distance between the current position and the first way points is calculated using the haversian formula as in (4) and if it is less than a certain threshold the queue is popped and the instruction is updated.When the queue is empty it means that the user has reached his destination.

$$p = \pi/180 \quad (1)$$

a = 0.5 - cos((lat2-lat1)*p)/2 + cos(lat1*p) * cos(lat2*p) * (1-cos((lon2-lon1)*p))/2 (2)

distance = 12742 * asin(sqrt(a)) (3)

The haversine formula determines the great-circle distance between two points on a sphere given their longitudes and latitudes. The formula in (3) is modified so that the response time of the system is fast and to make it computationally efficient. This is a special case of a more general formula in spherical trigonometry.

3. Modeling and analysis

YOLO

YOLO is short for You Only Look Once. It is a real-time object recognition system that can recognize multiple objects in a single frame. Object detection models are trained to look at an image and search for a subset of object classes. When found, these object classes are enclosed in a bounding box and their class is identified. Realtime is particularly important for object detection models that operate on video feeds, such as self-driving cars. The other advantage of real-time object detection models is that they are small and easy to wield by all developers.

The architecture of an object detector

All object detectors take an image in for input and compress features down through a convolutional neural network backbone. In image classification, these backbones are the end of the network and predictions can be made off of them. In object detection, multiple bounding boxes need to be drawn around images along with classification, so the feature layers of the convolutional backbone need to be mixed and held up in light of one another. The combination of backbone feature layers happens in the neck. It is also useful to split object detectors into two categories: one-stage detectors and two-stage detectors. Detection happens in the head. Two-stage detectors decouple the task of object localization and classification for each bounding box. One-stage detectors make the predictions for object localization and classification at the same time. YOLO is a one-stage detector, hence, You Only Look Once.

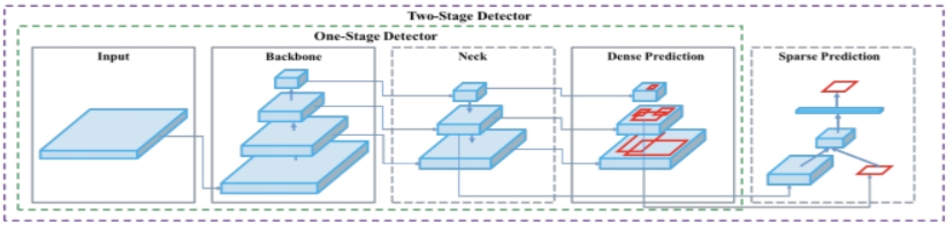

Figure 2: Object detector architecture[11]

DisNet

The architecture of the DisNet based distance detection is as shown in the figure 3. The camera image is input to the Object Classifier which is based on a state-of-the-art computer vision object detector YOLO (You Only Look Once) trained with COCO dataset . Its outputs are bounding boxes of detected objects in the image and labels of the classes detected objects belong to. The objects bounding boxes resulting from the YOLO object classification are then processed to calculate the features, bounding boxes parameters. Based on the input features, the trained DisNet gives as output the estimated distance of the object to the camera sensor.

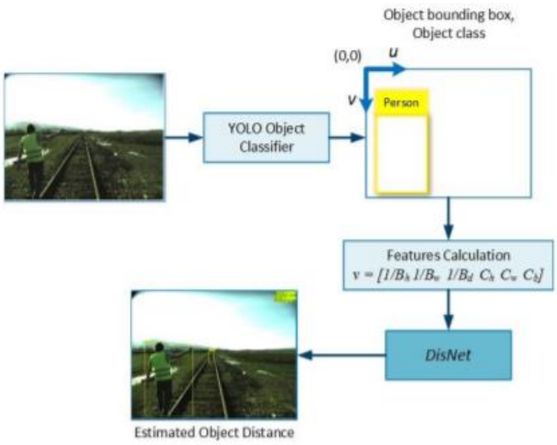

Figure 3: DisNet Model [12]

Street Navigation

To achieve street navigation, we need the following elements: location of the Destination, location of the user, information of the surroundings and finally being able to determine the best path to reach the destination. In Order to obtain the location of the user, the raspberry pi utilizes GPS. The location is continuously updated and so is the path. The tool has a voice assistant which can be communicated with whenever required by the user. If the visually impaired person needs to go from one place to another they can ask the voice assistant to navigate them to the place. The camera provides obstacle detection and the obstacle distance is measured through machine learning algorithms discussed later.

The fields inputted into the Navigation system are user request, user location and user orientation. Using these we can update the virtual map in real-time, the user can enter the destination with voice commands and the request will be sent to google maps API for navigation. The user will be guided to the destination by speech, the camera will help recognize any obstacle in the surroundings and alert the user if it is too close. This obstacle detection uses YOLO algorithm and to measure distance DisNet is used with it in combination.

The system will analyze and process the gathered information from the user's perspective in order to provide suitable guidance commands. These commands are updated in real-time based on the user's surroundings and location. The data required to achieve these commands are:

User's request: The desired destination of the user.

User's location and orientation: The GPS location of the user and what direction the user is facing.

User's perspective: The camera input which will see what the user is facing.

All the while the person moves the camera will assess the frame, detect obstacles using YOLO algorithm and estimate the distance of the obstacle with a combination of DisNet and YOLO . If the obstacle is close it will alert the user about the distance of the object and also its direction of approach.

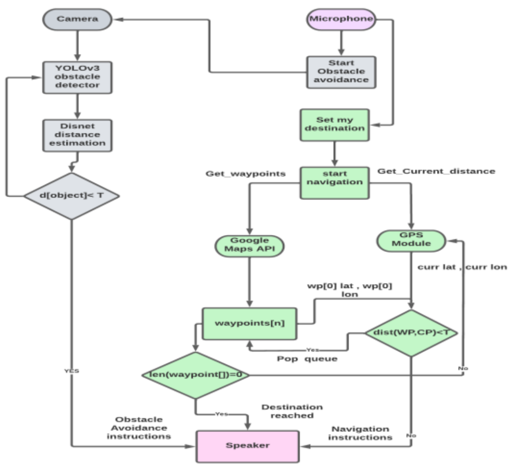

Figure 4: Logical design of the system

Figure 4 represents the logical view of the system where the distance estimation of the directions module and the user interface runs parallelly in the system. There are three threads that run simultaneously in the application. Main thread consists of the user interface, the YOLO DisNet cascade, and the navigation system. The user interface thread is responsible for receiving voice input from the user and processing their request according to the request it gives output or triggers other methods that need to be acted. The Yolo DisNet cascade is always acting when the device is on and constantly outputs any obstacles that are deemed to be dangerous. The navigation system thread works in conjunction with GPS to get current location and Google map interface to provide way points. The voice assistant will help the person navigate through voice messages from speakers. The multi-threaded environment has a shared mutex for the speaking engine so that only one thread is able to speak at a time. The mutex is implemented with a lock and free mechanism with proper priority queuing for good response time.

The proposed wearable device is a combination composed of hardware modules and three software modules to deal with the collected environmental information. The hardware module consists of a Pi-Camera, Raspberry Pi, speakers, microphone, and GPS module. The hardware are robust and compact which is helpful for the visually impaired person to use the system without it affecting his day-to-day work.

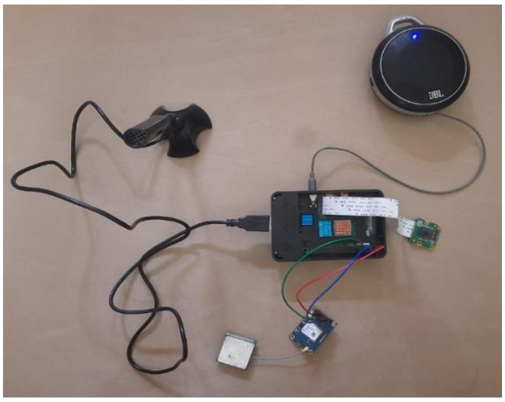

Figure 5: implementation of the system

The interfacing of the hardware module is done as shown in figure 5. The raspberry pi is given the power through the 1.5A, 5V power bank which is sufficient to drive the other hardware module

4. Results and discussion

The system is evaluated based on its performance, objective and response time in the real time environment the system is deployed and evaluated on the User-Interface system, object detection, and distance estimation, navigation directions,user-friendliness and response time of the system.

User-Interface

The output of various instructions based on the user input are shown in TABLE 1. The microphone is interfaced to the system and when the user gives the voice input into the system the system converts the voice input to the texts and processes it and the voice instructions are outputted using the speakers. The user-interface is multi programmed and will be running parallelly along with the obstacle detection and the navigational system hence the user can start, stop and restart the system through the voice instructions and the user can also make the system repeat its previous instruction as well.

TABLE I

AUDIO INSTRUCTIONS

Conditions	Audio Output
User says Set my destination	Where is your destination
User says How far is my destination	The destination is D km away
User says How long does it take to reach at the destination	The destination is T minutes away
User says Change my destination	Where do you want to change your destination
User says What is in front of me	Outputs the objects in the camera frame
Obstacle placed in front of the user with no feasible direction	Attention, an obstacle in front of you, turn left or right slowly
Obstacle placed on the left Attention	The obstacle is present on the left at a distance XX
Obstacle placed on the right Attention	The obstacle is present on the right at a distance YY

Object detection and Distance estimation

The Objection detection is implemented using the YOLOv3 algorithm and this algorithm was trained on the COCO dataset which contains photos of 91 objects types with a total of 2.5 million labelled instances in 328k images.The object detection runs when the user starts the system and starts detecting the object figure 6. if the user asks for the objects in the surrounding the system outputs the list of objects through the speakers.The detected object is later passed through the DisNet which calculates the distance between the object and the camera.ThisDisNet model takes in the output provided by the YOLO algorithm and based on it's class it estimates the distance figure 7. based on this distance the user can be provided with appropriate instructions to find the object or to avoid the obstacle.

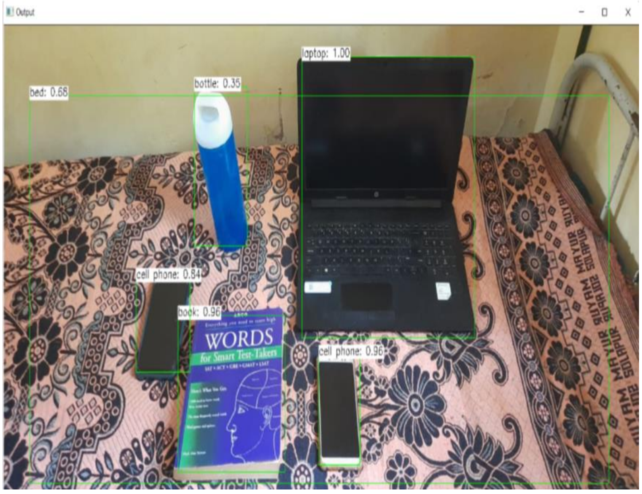

Figure 6: Object Detection

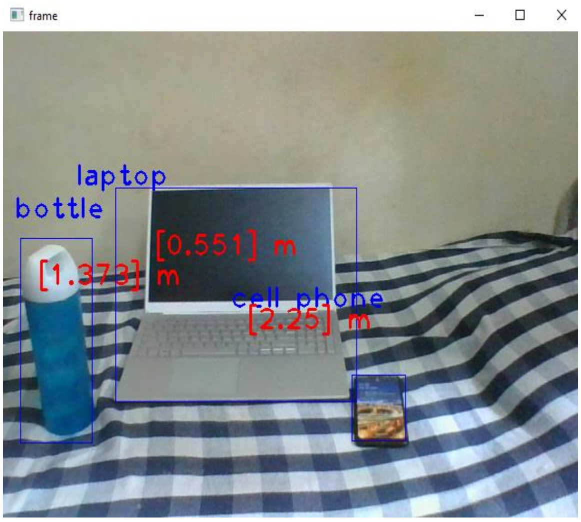

Figure 7: Distance estimation in DisNet

Navigation Directions

When the user sets his destination the request is sent to the Google maps API and sends the response. If a valid is found then the directions are provided to the user. In figure 7 is the Google maps response when the current location of the user is SJCE Mysuru and he sets his destination to BMH Mysuru we can see the ETA and the distance. These information is extracted from the response Json and formatted as a waypoint queue which consists of the instructions to reach for the particular waypoints as in figure 8 and the latitude and longitude of the waypoints as in figure 9 the user can update his destination at any given point of time and can refresh the by resetting his destination. When the user reaches the waypoint the queue pops up and the instruction is updated. This is continued until the queue is empty; the destination message is given to the user through the speakers when the user reaches his destination.

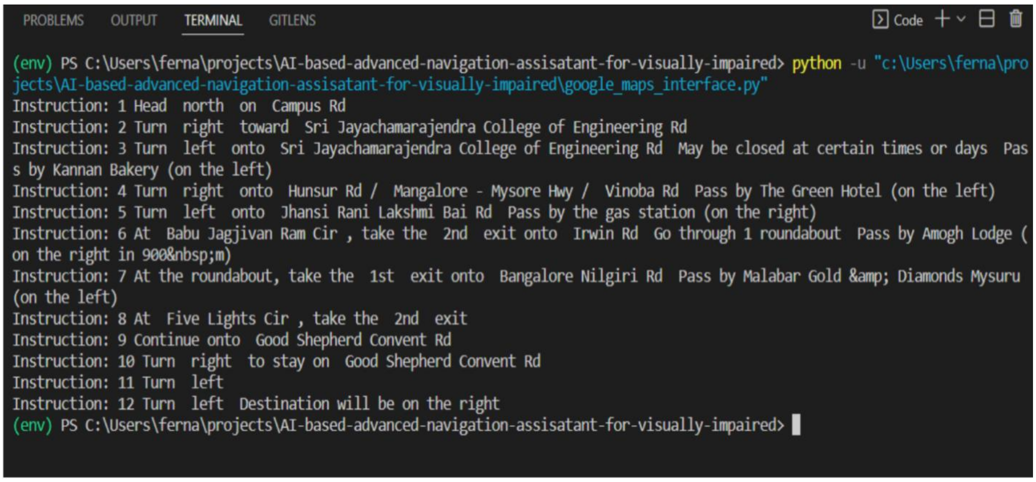

Figure 8: Output of the distance

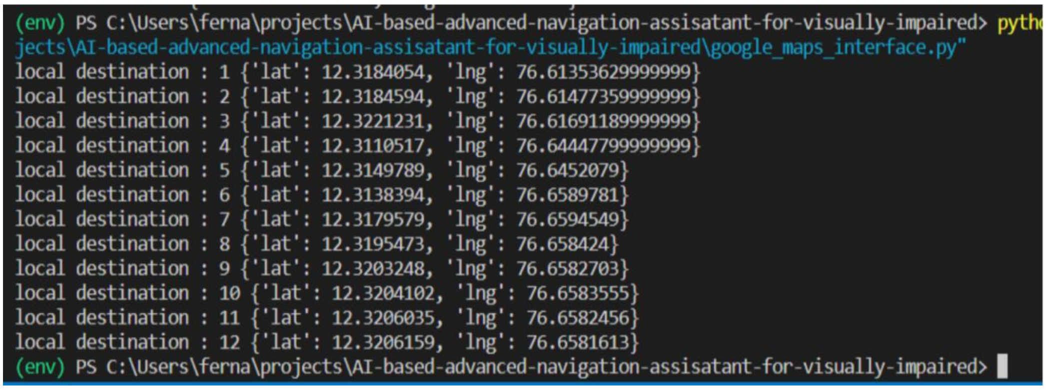

Figure 9: Output of the waypoints

Multiprogramming and Response time

The system to run properly must run the processes in parallel.The proposed system contains three major modules: The User-Interface, Object detection system and the navigation system. all these three modules are important and crucial in the navigation hence these modules must be run parallelly hence these three modules are multiprogrammed.

The visually impaired person can walk at his own pace hence the system must be built such that its response rate is greater than or equal to the speed at which the user walks. Hence each system must give a response within less than a second

TABLE II RESPONSE TIME

Processing Step	Average time
speech to text conversion	1.32 ms
regeneration of speech	1.73 ms
Time taken for all the modules to bootup	2.31 s

Object detection	3.51 ms
Disnet algorithm	2.98 ms
GPS module	6.21 ms
Frames per second processed by YOLO and Disnet	5 fps

The response time is calculated as an average taking 10 readings. the TABLE II represents the average response time taken by each processing step.

5. Conclusion

The project AI-based blind Assistant aids as an object recognition system and is a very helpful product for the blinds. This system will be designed comprising a web camera, a headset with a mic, where the control kernel is performed by Raspberry Pi. The project can identify the objects that are needed by the user. The user asks for the objects they need and if the object they asked is available in the camera range then the user is provided with the information about the object whether it is right, left, top, or bottom. The user will also get information about the location of the needed object. User can set the destination location he wants to reach and he is guided to navigate to that location safely avoiding any encounter with obstacles. Navigation is safe as distance from possible obstacles is determined using DisNet and object Detection models that will guide user to navigate safely. This is economical as well as an efficient device for visually impaired people. Thus, ultimately this could increase the confidence level of the user as well as make him/her feel secure. The device is compact and helpful to society.

References

[1] "Blindness and vision impairment" Who.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 28-Nov-2021].

[2] J. Faria, S. Lopes, H. Fernandes, P. Martins, and J. Barroso, "Electronic white cane for blind people navigation assistance," in 2010 World Automation Congress, 2010, pp. 1–7.

[3] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and technologies for blind and visually impaired navigation support: A review," IETE Tech. Rev., pp. 1–16, 2020.

[4] C. Ye and X. Qian, "3-D object recognition of a robotic navigation aid for the visually impaired," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 2, pp. 441–450, 2018.

[5] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Trans. Syst. Man Cybern. C Appl. Rev., vol. 40, no. 1, pp. 25–35, 2010.

[6] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Virtual-blind-road following-based wearable navigation device for blind people," IEEE trans. consum. electron., vol. 64, no. 1, pp. 136–143, 2018.

[7] A. Sen, K. Sen, and J. Das, “Ultrasonic blind sticks for completely blind people to avoid any kind of obstacles,” in 2018 IEEE SENSORS, 2018, pp. 1–4.
[8] M. A. Khan, P. Paul, M. Rashid, M. Hossain, and M. A. R. Ahad, “An AI-based visual aid with integrated reading assistant for the completely blind,” IEEE Trans. Hum. Mach. Syst., vol. 50, no. 6, pp. 507–517, 2020.
[9] B. Li et al., “Vision-based mobile indoor assistive navigation aid for blind people,” IEEE Trans. Mob. Comput., vol. 18, no. 3, pp. 702–714, 2019.
[10] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa, “NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment,” in Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility, 2017.
[11] M.-C. Kang, S.-H. Chae, J.-Y. Sun, J.-W. Yoo, and S.-J. Ko, “A novel obstacle detection method based on a deformable grid for the visually impaired,” IEEE transconsum. electron., vol. 61, no. 3, pp. 376–383, 2015.
[12] M. A. Haseeb, J. Guan, D. Ristić-Durrant, A. Gräser, and Member, “DisNet: A novel method for distance estimation from monocular camera,” Inria.fr. [Online]. Available: https://project.inria.fr/ppniv18/files/2018/10/paper22.pdf
[13] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, and M. Atri, “An evaluation of RetinaNet on indoor object detection for blind and visually impaired persons assistance navigation,” Neural Process. Lett., vol. 51, no. 3, pp. 2265–2279, 2020.
[14] H. Zeng, B. Yang, X. Wang, J. Liu, and D. Fu, “RGB-D object recognition using multi-modal deep neural network and DS evidence theory,” Sensors (Basel), vol. 19, no. 3, p. 529, 2019.
[15] R. Tapu, B. Mocanu, and T. Zaharia, “DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance,” Sensors (Basel), vol. 17, no. 11, p. 2473, 2017.
[16] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan, “A compressive sensing approach to describe indoor scenes for blind people,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 7, pp. 1246–1257, 2015.
[17] Chen, Xiangrong, and Alan L. Yuille. "A time-efficient cascade for real-time object detection: With applications for the visually impaired." 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops. IEEE, 2005.
[18] Wei, Yunchao, et al. "HCP: A flexible CNN framework for multi-label image classification." IEEE transactions on pattern analysis and machine intelligence 38.9 (2015): 1901-1907.
[19] Trabelsi, Rim, et al. "Indoor object recognition in RGBD images with complex-valued neural networks for visually-impaired people." Neurocomputing 330 (2019): 94-103.
[20] K. Yang, K. Wang, W. Hu, and J. Bai, ‘‘Expanding the detection of traversable area with RealSense for the visually impaired,’’ Sensors,vol. 16, no. 11, p. 1954, Nov. 2016.