Date of publication July 31, 2022, date of current version Aug. 12, 2022.

A Review on Vision-Based Obstacle Avoidance and Assistant Systems for Visually Impaired People

KLH Imesha [1#], G Gayamini [1], and B Hettige[2]

[1,2,3]1Department of Computer Engineering, General Sir John Kotelawala Defence University, Sri Lanka.
[1#]36-ce-0009@kdu.ac.lk

ABSTRACT Travelling is one of the biggest problems faced by the visually impaired community in their day today life. Even though the visually impaired people have their own methods for travelling such as with the aid of white canes and guide dogs, those traditional methods have lot of difficulties and i ssues. Because of those i ssues, researches has paid attention to develop assistive technologies which can help the visually impaired community for their day today travelling. Although the assistant technologies are developed in the past decades, the visually impaired community faces a lot of issues when using them because of the Disability Digital Devide. As a result of this, number of systems and technologies have been introduced to the word. Those systems use different technologies like vision modules, sensors and GIS maps but most of those systems are vision-based and they use different approaches to develop their vision modules such as deep learning, reinforcement learning, stereo imaging, enhanced image processing techniques, etc. So in this paper, we will focus on those vision-based systems to find the optimal way of developing a vison-based assistant system for visually impaired people by studying those technologies and by comparing and contrasting those technologies while understanding their used modules algorithms, efficiency, usability, functionalities and their advantages and disadvantages.

INDEX TERMS: Computer Vision, Deep Learning, Smart Wearable.

I INTRODUCTION

With the help of the development of new technologies, the world is focusing on using assistive technologies to help people with disabilities to accomplish their daily tasks easier and more efficiently in work, education, communication, traveling, entertainment and other day-to-day activities. These new technologies and devices have helped to enhance the quality of the lives of differently-abled people in many ways. Among those, assistive technologies that improve the mobility of blind or visually impaired persons were given specific attention.

Vision is considered as one of the most used and powerful sensing method human have. According to World Health Organization around 220 million people worldwide have a visual impairment. Among them, 40 million are blind and 180 million have low vision.[1]

Although the new technology makes our life much easier, there is still a divide between those who are not differently-abled and those who are differently-abled to use those technologies. This gap is known as Disability Digital Divide. This gap becomes larger when it comes to visually impaired people because most of the technological devices like computers and mobile phones are vision-based operating devices. Nowadays most of the assistive technologies are widely used in smartphones but visually impaired people face different kind of problems using smartphones such as it is hard to perform specific touchscreen gestures, unawareness of supported gestures in current app, getting lost while using an app and having no way to correct it, Unable to scroll through lists, slow screen reader, lack of knowledge about the meaning of specific sounds, slow text input, changing the written text is hard, unable to find the desired option, etc.[2] Also voice commanding and audio feedbacks have issues like difficulty of recognizing audio feedback and commands in noisy surroundings, The acoustic feedbacks are not enough to give relevant information, using voice commands in public is not good for visually impaired person's privacy.[3] Even in computer and web related applications, there are usability issues for visually impaired peoples such as they cannot interact with software interfaces, can't navigate from keyboard arrow keys, screen reading software compatibility issues, etc.[4]

Because of those described problems, researchers tried to develop assistive wearable devices which is specially design to give high usability to the visually impaired people. Among this devices, navigation and travelling assistant devices take a major place because travelling is one of the major problems facing by the visually impaired community. Generally, a Vision-Based Obstacle Avoiding System has three major components named Vision Input Module, The Processing Modules and the Feedback Module.

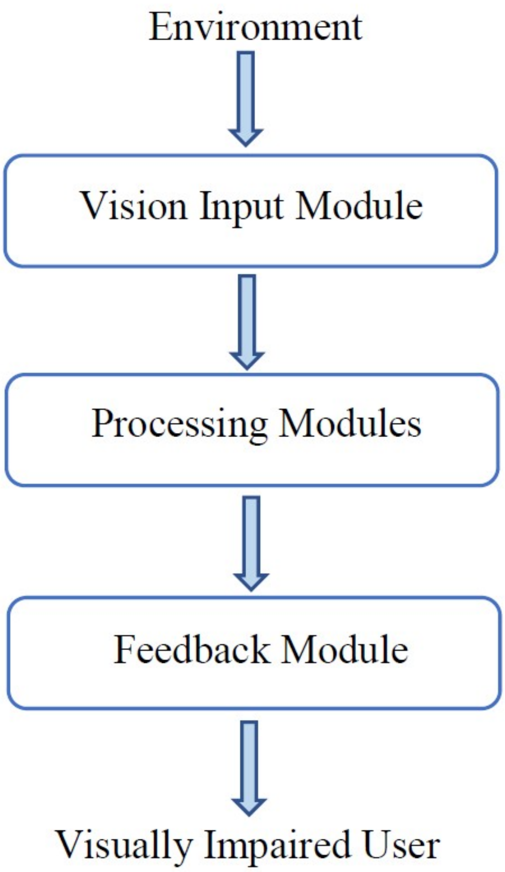

Figure 1. Flow of the process
Source: Author

he vision input module basically takes data from the environments as sequence of image frames or as sensory readings and inputs to the processing module. To take image frames as inputs, these systems use monocular cameras, stereo cameras and vision-related sensors like the Kinect sensor. For the sensory readings, most of the systems use IR sensors, Laser range finders, and Ultrasonic sensors.

Most of the time the output model is just a simple module that can give the relevant feedbacks to the visually impaired person. It generally have feedback modules like earphones and vibration modules.

The processing module can be a computer or a micro-controller which is responsible for the creation of point clouds, depth maps and the detection of edges, corners and other obstacles. This module is the most complex one in the system. Different systems use alternative methods and technologies to develop the processing module.

The designed and developed travel assistant devices use different technologies for the processing module according to their targeted functionalities. Most of the devices which were designed for indoor environments, uses short rage obstacle avoiding sensors like ultrasonic sensors, Radio-Frequency Identification (RFID) technology and models like ARIANNA (pAth Recognition for Indoor Assisted NavigatioN with Augmented perception) to identify different places, objects and paths in an indoor environment. But when analyzing the systems designed for the outdoor conditions, things get somewhat complicated. Most of these outdoor navigation systems have several modules. The main module is the vision module. Generally this module consists of a camera module. Stereo vision systems use 2 cameras (stereo cameras) and monocular vision systems use only one camera. Some of the systems also used depth sensors like Kinect sensors to identify and calculate the position of obstacles. These systems also use GPS modules to track the outdoor location and give some specific functionalities. The other main module is audio feedback module. Almost all the systems use audio feedback module because audio feedback is the easiest and the most suitable way to interact with visually impaired people. These feedbacks is mostly used to give warnings and also to give relevant instructions. In the upcoming chapter describe about those technologies and systems are described with more details to understand what are the pros and cons of those systems.

The rest of the paper organized as follows. Section II present background and related work in this field. Section III will summarize about the used technologies, algorithms and also compare and contrast them. Section IV will discuss about the conclusions we can made by comparing and analyzing different systems and section V will give the references.

II RELATED WORK

A number of researchers have studied about different technologies which is applicable to find a reliable solution to this problems facing by the visually impaired community. They have developed several systems which can give fair output using technologies like monocular and stereo vision with deep learning, reinforcement learning and other algorithms, RFID, ultrasonic technology, IR technologies, GPS, laser technologies, etc. But each of those systems have their own advantages and disadvantages in different sections like usability, reliability, speed, cost, etc. So, to get a clear idea about them it's better to review the systems developed by those different technological approaches.

A Deep Learning Approach

One of the popular technologies that used in these systems is the Deep learning approach. An interesting work done by St. Francis Institute of Technology[5] uses four-layered Convolutional Neural Network (CNN) which can scan the surroundings, detects and classifies the nearby objects as well as the distances to those relevant objects with a response time of 50ms. It also uses an ultrasonic sensor to increase the accuracy of the measured distance. This system uses audio feedback and vibrations to alert the VIPs. The work [6] provides technology with reusable way-finding with obstacle avoidance. To predict the navigation actions, they have trained a Recurrent Neural Network and for obstacle avoiding, they have fine-tuned a Convolutional Neural Network model. The paper [7] also shows how Convolutional Neural networks can be used to estimate the Transmission map for obstacle detection.

B Image Processing Techniques and Algorithms

Most of the systems use improved computer vision algorithms to Develop vision-based assistant systems for visually impaired people. The system [8] is an indoor-outdoor Navigation System specially designed for people who have low vision. This model uses ARIANNA (pAth Recognition for Indoor Assisted NavigatioN with Augmented perception) which is an indoor localization and navigations system. For outdoor conditions, it uses a computer vision system with several measurement sensors. The system has Geometry Based Path Identification which uses canny algorithms to detect path searching lines by image inputs. To improve the accuracy of those input images the system uses traditional image enhancement techniques like Gaussian smoothing and other enhancement filters to reduce noise effects. After the images convert into the canny form it can it process them to detect edges, lines and slopes. Color-Based Path identification is a functionality that make this system much better. Using this technology, the buildings like hospitals can apply color stripes on the floor and by detecting them the system can assist the visually impaired person. The system uses both the inertial sensors and the camera of the smartphone of the user as sensors

C Stereo Vision

When we consider about the systems which are using image processing techniques and algorithms, most of the researchers have paid attention for using stereo cameras because stereo cameras make it easy to calculate the depth and other details in the 3d space. The technics of these systems are mainly based on optical flow which can be described as the pattern of apparent motion of objects, surfaces, and edges in a visual picture induced by the relative motion between an observer.

The system in [9] developed in university of Alcala gives a reliable way to detect obstacles in outdoor environment and alert VIPs by audio feedbacks. As the first step, the ground plane detection is done using RANdom SAmple Consensus which is known as (RANSAC). Next the system does the Stereo Rig Calibration and Rectification to create the disparity map and calculate x, y and z coordination of obstacles. The intrinsic parameters and distortion parameters of each camera, as well as the extrinsic parameters (rotation, translation) between cameras, are estimated in the stereo rig calibration issue. Both cameras are calibrated independently and obtain the intrinsic calibration matrix for each of the cameras.

$$K = \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Where u_0 and v_0 is the principle point of the camera and f_x and f_y are the focal lengths. The rotation matrix between cameras RLR is the identity matrix, and the translation vector TLR encodes the baseline B of the rectified stereo rig. When (u_L, u_R, v) are the stereo image projections of the same point (because in stereo images $v_L = v_R =$ v), We can calculate X, Y and Z 3D point coordinates with respect to camera by using following equations.

$$Z = f.\frac{B}{(u_R - u_L)}$$

$$X = Z.\frac{(u_L - u_0)}{f}$$

$$Y = Z.\frac{(v - v_0)}{f}$$

After finding X, Y and Z we can find the distance and angle of the object by using the given equations.

$$Distance = \sqrt{X^2 + Z^2}$$

$$\alpha = \tan^{-1}(\frac{Z}{X})$$

After that the polar cumulative grid is projected onto it for counting about potential obstacles in the ground plane.

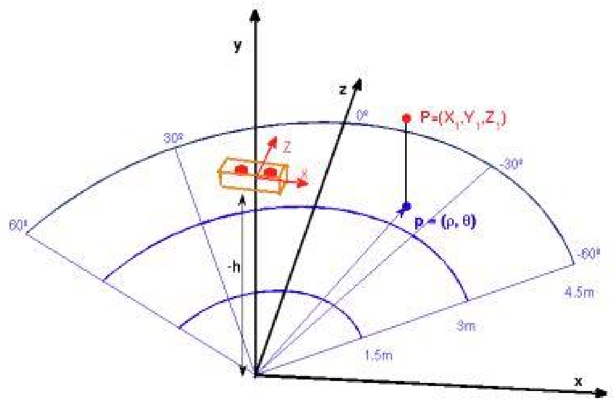

Figure 2. Creating the polar grid after calculation [10]

The system [11] also using the stereo imaging with object collision detection algorithm. It uses the disparity image or depth map as the previously discussed work[9]. That computer vision system is integrated with Blavigator prototype which is developed by University of Trás-os-Montes and Alto Douro (UTAD) known as a cheap easy to use mobile navigation system for visually impaired people providing ways to get to a given location and, while doing so, providing contextual information about obstacles and points-of-interest (POI) like zebra-crossings, building entrances by using GIS data and location data.

D Reinforcement Learning

Some researchers has tried to use monocular vision with reinforcement learning instead of stereo vision. In works like [12], they have proposed an algorithm that learns relative depth by monocular vision (only from single image) in outdoor environments. They have collected thousands of images with a laser range scan which gives the nearest obstacle in each direction of the image. Using that dataset, they trained the system with supervised learning algorithm and the system was able to accurately estimate the nearest object in the vision after the training.

E Navigation Robots

In the developed countries most of the visually impaired people use trained guide dogs in their day today travelling. Following that format, researchers have tried to develop assisting robots to guide the visually impaired people. The work like [13] and [14] uses robot assistants instead of smart wearables. These systems also has a stereo camera vision module. And they use some sensor modules like laser range finders and LiDAR in addition to the cameras to detect the obstacles. These systems use a tactical handle as the main feedback technology. It uses vibro-tactile feedback on the handle to convey directional information to the user.

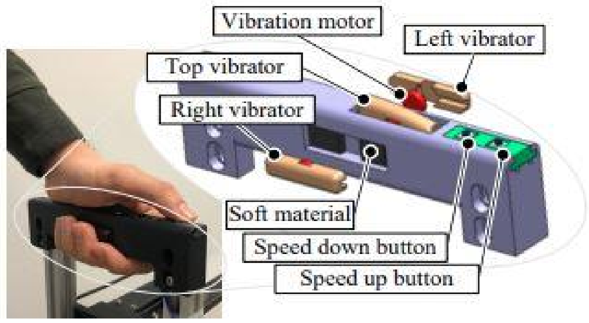

Figure 3. Modified suitcase handle with vibro-tactile feedback. [14]

F Vision and Depth Combined Sensors

When we study about the computer vision modules the most popular method is using cameras. However we can also use vision and depth combined sensors like Kinect sensor instead of cameras for special tasks like obstacle detection. Kinect sensor is Microsoft's motion sensor add-on for the Xbox 360 gaming console. It has an RGB color VGA video camera, a depth sensor, and a multi-array microphone. The camera detects the red, green, and blue color components as well as body-type and facial features.

The system [15], [16] and [17] are good examples for the systems which are using Kinect sensor and depth sensors for obstacle avoiding smart wearables. The obstacle avoiding of this system work in several steps. The point cloud registration stage seeks to create point cloud using information from the Kinect sensor like color, depth, and accelerometer data. To create the 3D point cloud from Kinect data, x, y and z coordination are calculated using this equations.

$$P3D_x = \frac{(x_c - cx_c) \times depth(x_c, y_c)}{fx_c}$$

$$P3D_y = \frac{(y_c - cy_c) \times depth(x_c, y_c)}{fy_c}$$

$$P3D_y = depth(x_c, y_c)$$

where xc and yc is the pixel coordinate in color image, cxc, cyc, f xc, f yc is taken from color intrinsic matrix, depth(xc, yc)is the depth value of pixel.

As the next step, plane segmentation is done to determine the dominant places from the point cloud by using the plane segmentation method proposed in [18] that permits real-time segmentation of point cloud data into various planes.

After processing plane segmentation step, Ground and wall detection is done as the last step.

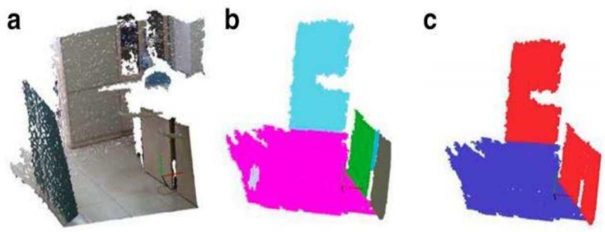

Figure 4. Plane segmentation and ground and wall detection results: a point cloud; b segmented planes; c detected ground (in blue) and wall planes (in red)[15]

The work [16] is also using the Kinect sensor but in addition to creating the depth map it uses typical image feature detecting algorithms like Sobel edge detector to detect edges clearly, and Harris & Stephens detector to detect the corners of the objects in the vision and finally blobs that shows the optical flow is detected using the SIFT detector.

III DISCUSSION

This section of the paper provides a brief discussion on the works, there pros and cons, suggestions and challenges we studied in early chapter. First it is better to summarize about the different approaches for the vision based obstacle avoidance and assistant systems because it is easy to compare and contrast to identify their plus and weak points after summarizing all the systems. In the following table we will categorized the system according to the technological approaches they used and the advantages and disadvantages of those approaches.

When comparing the different technical approaches mentioned above, they have their own advantages and disadvantages. A highly acceptable system should consider about technical factors or functional requirements like reaction time, accuracy, reliability as well as the non-functional requirements like usability, mobility, cost, etc. But achieving this requirements is a challenge due to many reasons.

A Cost

Generally, the visually impaired community is lower paid than others so the cost of the device is a critical fact for them. The vision module (camera or sensors), the processing unit (computer or microcontroller) and, the power supply will be the main components that affect the final cost. By using technologies that use lower computing power, we can complete the system with a low-power processing unit. Also, it is important to use technologies that use low-cost vision modules (unlike costly vision modules like stereo cameras).

B Size and Mobility

Since these systems are traveling assistant systems, they have to be small in size with higher portability and it is better if the system could be developed as a wearable device. But the challenge is most of the technological approaches need high computing power and need to operate in real-time so most of the times it is necessary to move for larger computers or microcontrollers and that will be a challenge to make the system smaller. The other main factor is since we need to operate these systems in outdoor traveling for a longer time we need to add larger batteries for more battery capacity.

C Require Higher Accuracy

The result of an error in this kind of system can be critical. So the system must provide very high accuracy and reliability to the user.

So we need to find the best approach to achieve the given requirements while facing those discussed challenges. When we consider about the physical design of the systems, smart wearables with low weight is always gives more mobility and for any kind of place. But the heavy systems like navigation robots can't be used in places like staircases. And when consider about the sensor and vision modules single camera module are cheaper than stereo camera modules and for distance measuring, sensors with laser technologies always give more accurate outputs and more range but higher on cost. So we need to select the sensors carefully according to our needs so we can develop a high reliable, accurate assistant device with high usability to address the day today travelling problem of visually impaired people.

The following table will facilitate you to compare the relatable have a clear idea about the relatable technologies, used module and the advantages and disadvantages of each technological approach.

IV CONCLUSION

Travelling is one of the main problems facing by the visually impaired community in their day to day lives. Although lot of assistive technologies are available, most of them are not user-friendly for visually impaired people because of the Disability Digital Divide gap. It is hard to use mobile applications and web based applications specially for visually impaired people. So smart wearable devices has to come forward to help them in their day today travelling and develop the navigation capabilities. Vision based obstacle avoidance and assistance technology is the most popular approach to address this problem. The related works shows us it is important to use technologies with high

Table 1. Monocular vision with machine learning other algorithms

Technological Approach	Description	Modules Used	Advantages	Disadvan-tages	Refere-nces
Deep learning models	Deep learning approach can give accurate localization and navigation instructions. Recurrent Neural Network and Convolutional Neural Network is used in training and recognition process.	Camera, Ultrasonic Sensor	Low reaction time(50ms), High Accuracy, Low cost	Complex training process	[5] [6] [7]
Image processing techniques and other algorithms	By using the image processing techniques which identify edges, lines, corners and identifying obstacles using them.	Monocular Camera	Low cost	Low accuracy when comparing with other techniques	[8]
Reinforcement learning	Collect thousands of images with a laser range scan which gives the nearest obstacle in each direction of the image. Using that dataset, train the system with a supervised learning algorithm. It gets more accurate with larger datasets.	Monocular Camera	High Accuracy, Low cost	Need very large data set for higher accuracy	[12]

Table 2. Special Camera Technologies and Sensors

Technological Approach	Description	Modules Used	Advantages	Disadvan-tages	Refere-nces
Stereo imaging	Calculating 3D coordination using two images captured at the location buy from two cameras. And create the disparity map and estimate the ground plane for detect the obstacles.	Stereo Cameras	Not too complex, Low cost	High cost for the vision module (stereo cameras)	[9] [11]
Cabot, Navigation robots	Robots (not wearable) developed with main vision module with many sensor modules for scan and identify the obstacles.	Cameras, Ultrasonic sensors, LiDAR, Laser technologies	High reliability	High Cost, Less mobility	[13] [14]
Vision and Depth sensors	These systems are designed with depth and vision combined sensors like Kinect sensor which make it easy to calculate x, y, z coordinates and optical flow for detect obstacles edges, corners and ground plane.	Kinect sensor, Depth sensors	Low reaction time	Sensitive to infrared and highly reflective objects.	[15] [16] [17]

accuracy and reaction time but also need to have high usability, mobility and low cost. The computer vision systems with deep learning, reinforcement learning, enhanced image processing techniques with relevant cameras or sensors have given acceptable solutions for the addressed problem in the related works. With adding relevant upgrades with the newly updated technologies to those systems, they will give fair service for visually impaired community to make their day today travelling easier and efficient.

REFERENCES

[1] World Health Organization, "World report on vision," p. 180, 2019.

[2] A. Rodrigues, H. Nicolau, K. Montague, J. Guerreiro, and T. Guerreiro, "Open Challenges of Blind People using Smartphones," ArXiv190909078 Cs, Oct. 2019, Accessed: Oct. 24, 2021. [Online]. Available: http://arxiv.org/abs/1909.09078.

[3] H. H. Qureshi and D. H.-T. Wong, "Problems Facing by Visually Impaired People during Interaction with Mobile Applications," p. 8, 2019.

[4] N Wedasinghe, N Sirisoma, APR Wicramarachchi, "Web Mobile and Computer Accessibility: Issues Faced by Sri Lankan Visually Impaired Community," 11TH Int. Res. Conf., Sep. 2018.

[5] D. K. Yadav, S. Mookherji, J. Gomes, and S. Patil, "Intelligent Navigation System for the Visually Impaired - A Deep Learning Approach," in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, Mar. 2020, pp. 652–659. doi: 10.1109/ICCMC48092.2020.ICCMC-000121.

[6] F. Ahmed, M. S. Mahmud, and M. Yeasin, "RNN and CNN for Way-Finding and Obstacle Avoidance for Visually Impaired," in 2019 2nd International Conference on Data Intelligence and Security (ICDIS), South Padre Island, TX, USA, Jun. 2019, pp. 225–228. doi: 10.1109/ICDIS.2019.00041.

[7] J. O. Gaya, L. T. Goncalves, A. C. Duarte, B. Zanchetta, P. Drews, and S. S. C. Botelho, "Vision-Based Obstacle Avoidance Using Deep Learning," in 2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR), Recife, Brazil, Oct. 2016, pp. 7–12. doi: 10.1109/LARS-SBR.2016.9.

[8] D. Croce et al., "An Indoor and Outdoor Navigation System for Visually Impaired People," IEEE Access, vol. 7, pp. 170406–170418, 2019, doi: 10.1109/ACCESS.2019.2955046.

[9] A. Rodríguez, L. M. Bergasa, P. F. Alcantarilla, J. Yebes, and A. Cela, "Obstacle Avoidance System for Assisting Visually Impaired People," p. 7, 2012.

[10] A. Rodríguez, J. J. Yebes, P. Alcantarilla, L. Bergasa, J. Almazán, and A. Cela, "Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback," Sensors, vol. 12, no. 12, pp. 17476–17496, Dec. 2012, doi: 10.3390/s121217476.

[11] P. Costa, H. Fernandes, P. Martins, J. Barroso, and L. J. Hadjileontiadis, "Obstacle Detection using Stereo Imaging to Assist the Navigation of Visually Impaired People," Procedia Comput. Sci., vol. 14, pp. 83–93, 2012, doi: 10.1016/j.procs.2012.10.010.

[12] Jess Michels, Ashutosh Saxena, Andrew Y. Ng, "High Speed Obstacle Avoidance using Monocular Vision and Reinforcement Learning," 22nd Int. Conf. Mach. Learn. Bonn Ger., 2005.

[13] G. Capi and H. Toda, "Development of a New Robotic System for Assisting Visually Impaired People," Int. J. Soc. Robot., vol. 4, no. S1, pp. 33–38, Nov. 2012, doi: 10.1007/s12369-011-0103-1.

[14] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa, "CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People," in The 21st International ACM SIGACCESS Conference on Computers and Accessibility, Pittsburgh PA USA, Oct. 2019, pp. 68–82. doi: 10.1145/3308561.3353771.

[15] V.-N. Hoang, T.-H. Nguyen, T.-L. Le, T.-H. Tran, T.-P. Vuong, and N. Vuillerme, "Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile Kinect," Vietnam J. Comput. Sci., vol. 4, no. 2, pp. 71–83, May 2017, doi: 10.1007/s40595-016-0075-z.

[16] N. Kanwal, E. Bostanci, K. Currie, and A. F. Clark, "A Navigation System for the Visually Impaired: A Fusion of Vision and Depth Sensor," Appl. Bionics Biomech., vol. 2015, pp. 1–16, 2015, doi: 10.1155/2015/479857.

[17] K. Yelamarthi and K. Laubhan, "Navigation assistive system for the blind using a portable depth sensor," in 2015 IEEE International Conference on Electro/Information Technology (EIT), Dekalb, IL, USA, May 2015, pp. 112–116. doi: 10.1109/EIT.2015.7293328.

[18] D. Holz, S. Holzer, R. B. Rusu, and S. Behnke, "Real-Time Plane Segmentation Using RGB-D Cameras," in RoboCup 2011: Robot Soccer World Cup XV, vol. 7416, T. Röfer, N. M. Mayer, J. Savage, and U. Saranlı, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 306–317. doi: 10.1007/978-3-642-32060-6_26.

ACKNOWLEDGMENT

This review would not have been possible without the great support given by my two supervisors Mrs. G Shanmuganathan, and Dr. B Hettige by educating and correcting me in every step to the final paper. And my grateful thanks goes to Faculty of Computing, General Sir John Kotelawala Defence University for giving me the learning opportunities and facilitating me to collect all the resources.

AUTHOR BIOGRAPHIES

KLH Imesha is a 4th year Computer Engineering undergraduate in Department of Computer Engineering, General Sir John Kotelawala Defence University, Sri Lanka. The author is interested in Computer Vision, Deep Learning and Embedded Systems and the related fields.

Gayamini Gnanasuganthan working as a Lecturer (Prob.) in the Department of Computer Engineering, Faculty of Computing, General Sir John Kotelawala Defence University. This author was awarded the M.Sc. degree in Computer Science and Technology from Wuhan University of Technology, Wuhan, P.R. China in the year 2020. Her current research interests lie in Artificial Intelligence, Multi-agent systems, Machine Learning, IoT and Robotics & Automation.

B Hettige is Head of the Depart-ment of Computer Engineering, Fac-ulty of Computing, General Sir John Kotelawala Defence University. His research interests include Multi-Agent Systems, Machine Translation, Sinhala Language and Computational Grammar.