An automated navigation system for blind people

Md. Atiqur Rahman, Sadia Siddika, Md. Abdullah Al-Baky, Md. Jueal Mia
Department of Computer Science and Engineering, Daffodil International University-1207, Dhaka, Bangladesh

Article Info

ABSTRACT

Proper navigation and detailed perception in familiar or unfamiliar environments are the main roles for human life. Eyesight sense helps humans to abstain from all kinds of dangers and navigate to indoor and outdoor environments. These are challenging activities for blind people in all environments. Many assistive tools have been developed by the blessing of technology like braille compasses and white canes that help them to navigate around in the environment. A vision and cloud-based navigation system for the visually impaired or blind person was developed. Our aim was not only to navigate them but also to perceive the environment in as much detail as a normal person. The proposed system includes ultrasonic sensors detecting obstacles, stereo camera to capture videos to perceive the environment using deep learning algorithms. Face recognition approach identified known faces in front of him. Blind people interacted with the whole system through a speech recognition module and all the information was stored in the cloud. Web and android applications were developed to track blinds so that guardians were monitoring them while visiting and reached them in an emergency. The experimental results showed the proposed system could provide more plenty information and user-friendly interaction.

Corresponding Author:

Md. Atiqur Rahman
Department of Computer Science and Engineering, Daffodil International University
Dhaka-1207, Bangladesh
Email: atiqur15-9866@diu.edu.bd

1. INTRODUCTION

In our daily life moving from one place to another place is a common term and easy to navigate anywhere. But this is not easy for a visually impaired person, especially for blind people. We asked many blind people about their daily life and how they navigate from one place to another then they answered while navigating they face many problems. They don't know where they are, which obstacle in front of them, what is the objects, also face no information about the objects and another important thing is that when they want to know something, they have to ask another person to get the information. Statistics show that blind and visually impaired person faces navigation problem in their daily activities.

The statistics of the world health organization (WHO) shows, about 385 million people have visual impairments, among them about 39 million are blind in which 82% are blind whose age is 50 or older [1]. According to WHO (2012) around 4.24% are visually wretched and 0.58% are blind (estimated)-best-corrected visual acuity (BCVA) 0.05% in the better eye and 3.65% have low vision. Overall, among 39 million, 15% from Africa and gradually 2.7 million (7%), 23 million (67%), 3.2 million (8%), 5 million (12.5%) from Europe, Asia, America, and Eastern Mediterranean (re-2015). And among them, 82% of them are over 50 years and in Bangladesh, ages 30 and above is 750,000, 85 percent of them become blind due to cataracts [2]. Nowadays It is becoming an important term to develop a system for blind people to navigate as normal people.

In this research, an internet of things (IoT-based) new navigation system for blind people is designed. This will not only give guidance but also provide perceiving the environment as much as possible like normal people, such as which object is present in front of the blind (e.g., person, car, bus, currency, and chair), his current location, and recognizing known faces in front of. Also, the blind can get the distance of any person within 5 m and to know something from the internet saying his query. In this system, we have used raspberry pi to implement a navigation system. A five-megapixel stereo camera is used as a virtual eye for the blind, an ultrasonic distance sensor is used for obstacles detection in front of him and to measure the distance from the blind person. The proposed system captures the video sequence from the environment through the camera then processes the query through the system with voice delivered by the blind. While navigating in the outdoor global positioning system (GPS) module, it takes the location info and sends it to the cloud. Then the cloud stores these data and provides the current location to his guide. The system also uses the cloud for storing different types of data while blind interacting with the devices (e.g., search query and location data). The system uses a smartphone hotspot with a 4G mobile data connection to transfer data into the cloud and search web queries. The android application tracks the blind while visiting the outdoor environment. The web application helps the guidance to track a blind person from anywhere. This web application provides an overview of the system, visited the area, and the last location of their blind. The speech recognition module provides smart interaction with the system for the blind that can easily communicate through their formal English language. In the bellow section, we have discussed existing work which implements different technologies.

Many navigation systems already exist. But they work with some limited datasets and those can perceive a little information. Some of them are presented that helps to differ from other works. Dunai *et al.* [3] proposed a navigation system using complementary metal oxide semiconductor (CMOS) time-of-flight sensors, some are using radio frequency identification (RFID). These elements or systems are generally costly and CMOS time-of-fight sensor-based navigation system is wired, which is not user-friendly. Also, RFID-based navigation systems have a chance to lose data to read data at a time. Bai *et al.* [4] introduce a cloud and vision-based navigation where they proposed a system that includes speech recognition, simultaneous localization, and mapping, path-planning, using a deep learning approach which can communicate through voice. Bornschein *et al.* [5] stated a model that two-dimensional tactile pin-matrix display-based navigation system which has many limitations like data limitation and they worked on different kinds of input modalities, such as palettes with standard shapes or gesture interaction, freehand drawing by a digitizer stylus, and enabling to blind to create a drawing. Sivan and Darsan [6] ambient assisted living, mobile technology, ultra-sound systems are organized into different parts for collecting data then organizing them, then detecting the data and then ultrasound system converts those into voice and transfers the results to mobile phone. Setiadi *et al.* [7], Sarakhman *et al.* [8] states that using neural networks to help blind people. The researchers used two cameras to detect the pedestrian path and light detection ranging (LiDAR) to detect the surroundings. At first, the model can take an image by the camera then they get 3 voice attitude information. Majid *et al.* [9] stated a model that some machine learning-based systems are developed by convolutional neural network (CNN) algorithm and deep learning for object detection and categorized them where the system consists Arduino UNO, ultrasonic sensor and processing unit and you only look once (YOLO) is used for image segmentation and classification. Vijlyakumar *et al.* [10] stated a model that navigates blind people that guide them about objects as provides the distance of the object and it also provides the audio jack to insist them with object information and they have used single shot detection (SSD) algorithm to detect an object and find distance monodepth algorithm. Abirami *et al.* [11] stated a model that navigates and implements a system that is based on a robotic system. It's actually based on a voice navigation system that is commonly used and they used an assistive robot, blind pilot, which guided blind users. Blind pilot, and red, green, and blue (RGB) camera, presents the position of the object and uses lidar to build a 2D map of the surroundings. Ismail *et al.* [12] presents an effective solution based on speech recognition for patients, elderly people and disabled people with low-cost easy control system IoT devices and to enhance their system they used support vector machine (SVM) and dynamic time warping (DTW) algorithm. Compared to the existing technologies we have proposed a smart navigation for both indoor and outdoor environments with different machine learning approaches and ensures safe navigation, more feasible and effective than the present navigation system with the branch of advantages.

2.   RESEARCH METHOD

Though this is an IoT-based navigation system that uses several hardware components. All the sensors are implemented in Raspberry pi 4 model B. In this part, we describe implementation methodologies to develop our navigation system. Sensors are used to get the environmental details that help to build a system with proper navigation and all the equipment are available on the market that is used in this proposed

system. The system can be compared to a human brain, guides blind where they can visit with perceiving the surrounding. Overall, it's assured blinds security, this raised method initiates emergencies handling, also blinds can communicate with their guides in a complex situation. The web and android applications are designed to support and track the blind. The masonry of the raised system is given below. Speech processing, distance measurement, optical character recognition, and object and currency detection is the major part of the proposed system. The Figure 1 shows the architecture of our developed system. When a user speaks something to the system then specific module will be invoked with respect to user's query. Such as user want know something about 'Tell me about amazon' then the search module is invoked and search from the web and gives the result through voice, if the user wants to know the person or object distance then our system will tell that user what was the distance of object from him or her, user also able to get the currency, object and face identiy through detection and recognition module. We store the GPS (latitude, logitude) data into cloud that helps the guide to track their visits using web and Android applications. In the following paragraphs, this proposed methodology is described.
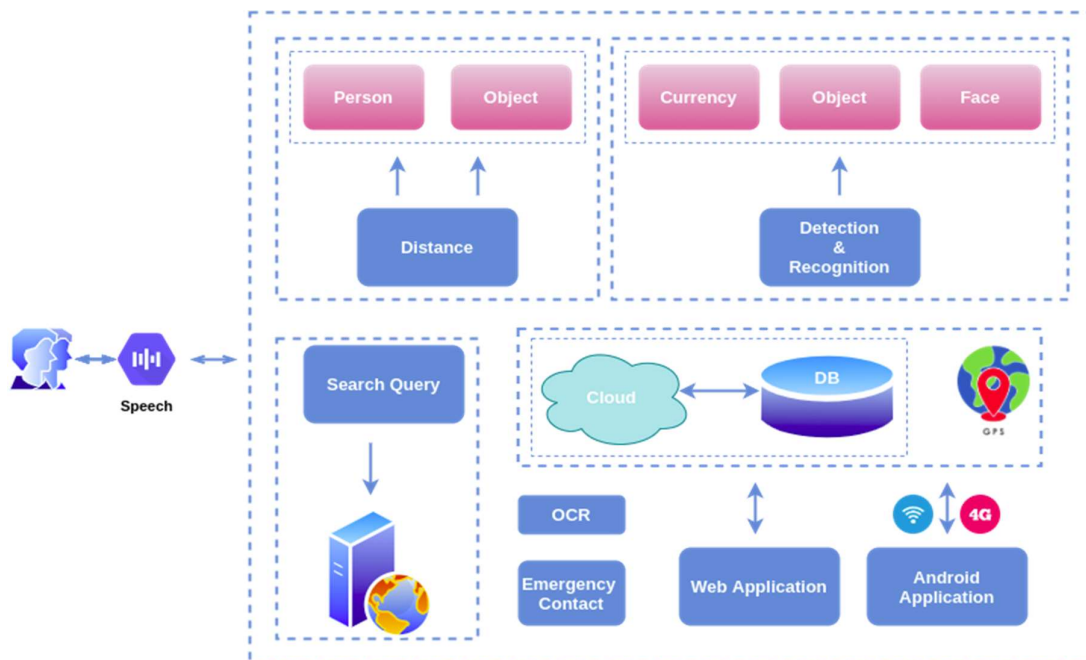

Figure 1. The main architecture of our proposed navigation system for the blind

2.1. Distance measure

The distance part is divided into two parts one is measuring distance using an ultrasonic sensor and another one is face distance or person distance using a camera attached with the device. The following two distance parts are described.

2.1.1. Distance measurement using ultrasonic sensor

To detect obstacles, the system used an ultrasonic sonic sensor. In ultrasonic (hc-sr04) consists of two transducers where one works as a converter that transmits electrical to ultrasonic sound pulses and another one is a receiver that listens for the transmitted pulses and the trigger pin is triggered when the pulse is at least 10 µS (10 microseconds) [13]. When the trigger pin is high then the sensor is emitted by an eight-cycle sonic burst at 40 kz and then it sets the echo pin to high till the sonic burst returns, it is called object reflection, and that the pulse length is proportional to the distance between the object and sensor. Figure 2 shows the illustration of the effects of sound where trigger pin (upper in Figure 2) transmits waves to detect objects, if found then echo pin (lower on in Figure 2) receives the return wave.

To measure the distance from the object using two pulses and the speed of the sound is 340 meters per second. In (1) states the speed S of distance D with Time T:

$$S = V * T \tag{1}$$

From 2 cm to 450 cm the ultrasonic provides accurate and stable measurements which are focused fewer than 15 degrees with an accuracy of about 2 mm.
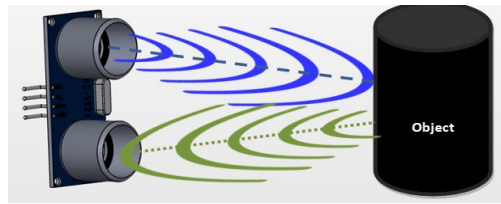

Figure 2. Sending and receiving of ultrasound wave pattern of ultrasonic distance sensor

2.1.2. Person distance measurement

In order to measure the distance from the person, utilize triangle similarity and the triangle similarity is having a known width W of an object and in general, the width of the known face is about 14 cm. The system captures the person's image through the camera and measures the discernible width in pixel P and helps to derive the perceived length F of the camera. The equation we have followed:

$$F = (P * D)/W \tag{2}$$

But when the stereo camera is moving in both nearest and further away from the object, the system applies triangle similarity to get the object distance to the camera, and then our desired equation is:

$$D' = (W * F)/P \tag{3}$$

In this system, the camera is used complementary metal oxide semiconductor (CMOS) sensor to run a distance measuring algorithm and the camera used in this system acts as a virtual eye for a blind person. The measuring algorithm is described below. The distance between CMOS and the distance of the lens is called focal length f. At initial, the distance between lens and object is d and the height of an object is h. The height of this object in the CMOS sensor of the angle of $\Theta1$. When the object comes closer to the camera lens the new height of the object is b in CMOS of angle $\Theta2$ and the distance between object and camera will be d-m. After the above algorithm, the final equation of distance is:

$$d = m/(1 - a/b) \tag{4}$$

Figure 3 shows the mechanism of how the above algorithm is used to measure distance using a CMOS image sensor. The camera attached to this system takes the images dynamically and gives the distance a person from the camera following the above algorithm.
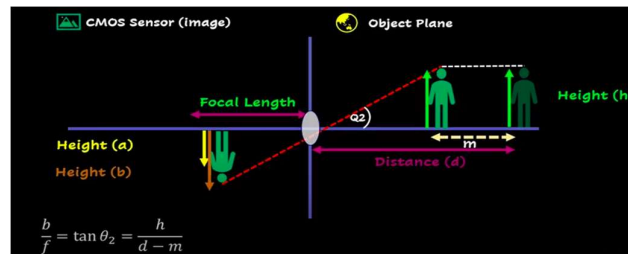

Figure 3. How the CMOS sensor works for measuring the focal length from the camera to person

2.2. Detection and recognition

In order to provide an adequate navigation guide for the blind, the system includes detection and recognition module and this section is classified into three parts as object detection and recognition, currency detection and recognition, and face recognition. To detect objects and currency, the system uses YOLO

framework neural networks which are used to get real-time object detection. The reason behind using the YOLO algorithm is because of its speed, higher accuracy, and learning capabilities that are high [14]. There are many algorithms present to detect objects but, we have used yolo because of its speed and higher accuracy. Figure 4 shows the one-stage detecting algorithm because two-stage is slower than one stage. The algorithm for detecting and recognizing objects is described below. Let's briefly discuss how yolo detect and identify objects.

Residual blocks-the image is split into various grids first which have a dimension of *m x n*. Every grid is responsible to detect objects when the objects appear in this grid. Bounding box regression-bounding box highlights of an object outline and these bounding boxes have a height (bh), width (bw), the center of the bounding box (bx and by), and the class name by letter c (e.g., person, car, and cup). Figure 4 shows the yellow outline that represents the bounding box details y and it uses single bounding box regression.

$$y = (pc, bx, by, bh, bw, c) \tag{5}$$

Intersection over union-describes how the boxes are overlapped and yolo is used intersection over union to provide a perfect outbox of object surroundings properly. Each cell is also responsible to predict bounding boxes and confidence scores. Combining the above three techniques are applied to predict the objects on the images and shows the confidence score with higher accuracy. Figure 5 shows the main mechanism behind the Yolo algorithm that is used in this system to identify several objects. At first, the system takes an image as input, then applies one stage detector to extract the features, bounding boxes, and then the dense prediction is used to predict the objects.
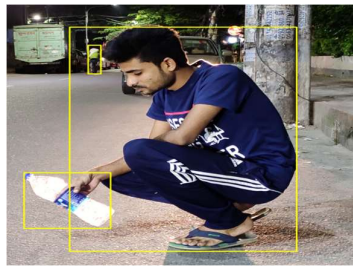

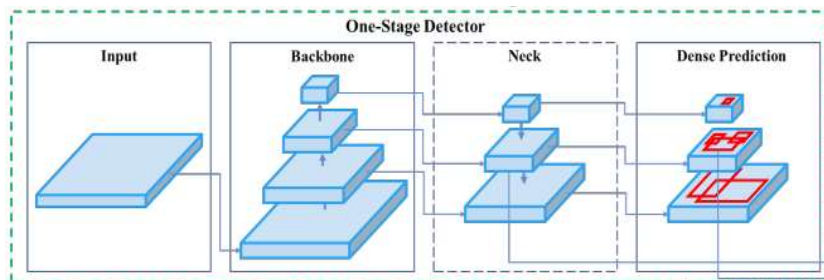Figure 4. Draw the bounding boxes if the trained objects are found

Figure 5. One stage object detection architecture of Yolov4

2.2.1. Object detection and recognition

In order to navigate perceiving the object's information, the system uses object detection and recognition techniques. The system takes the image from the environment then applies the object detection and recognition algorithm to classify and identify which object this is. This is section is divided into two parts, indoor and outdoor object identification in real-time. The system follows architecture that uses darknet as a backbone to identify objects. Furthermore, compared to other deep learning algorithms yolo has good scalability. The proposed system uses COCO dataset that has 330 k$^+$ images and 270 k$^+$ labeled images [15].

2.2.2. Currency detection and recognition

In order to identify currency, this system is used a Bangladeshi currency dataset. We could not find full datasets on the web for that most of the images are manually captured by our smartphone at various

backgrounds and different perspectives and the rest of the images are collected from the websites. Including 350 images these currency datasets are established by ourselves and are classified into 10 classes and the class refers to ৳১, ৳২, ৳৫, ৳১০, ৳২০, ৳৫০, ৳১০০, ৳২০০, ৳৫০০, ৳১০০০.

2.3. Face identification

The face recognition approach helps to navigate blind people in a more advanced way and we have used computer vision technology to determine locations and sizes of human faces in arbitrary (digital) images. In this technology. it detects faces by ignoring anything else, such as trees, bodies, or any parts of images. Figure 6 shows the diagram of how the system identifies faces. The system takes the image and then finds the faces in the captured image by locating where the faces are present, after locating faces it extracts the features. Comparing trained images, it makes a prediction. Using the above methods, the system can identify the face no matter if the face is moving or not then encode the face data to get better accuracy with 128 measurements using the trained network [16].

We know that the human can easily detect know faces though the faces are are moving or not but our computer looks totally different. To overcome this situation, we have used **Face landmark Estimation** algorithm that come up with 68 specific points called landmarks which is shown on Figure 7 that exist on every face (top top othe chin, outside edge edge of each eye and inner edge eyebrow).
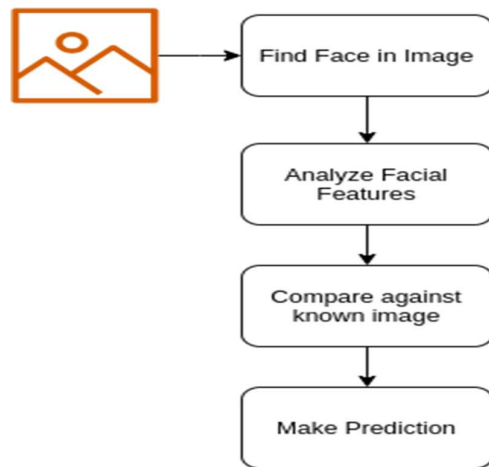
Figure 6. Diagram of the procedure of how face recognition works

Figure 7. The 68 landmarks that every face contains to uniquely identified faces

2.4. Optical character recognition

Electrical or mechanical conversation of typed, handwritten, or printed text into machine-encoded text is done by optical character recognition (OCR) whether a document is scanned or a photo document or a scene-photo or subtitle text are superimposed on an image. In this article, we have used OCR to recognize text from images or roadside directions or text that provides blind to read the documents or text from the images or pdf. The system takes the requested image first. Though it takes a colored image so that we need to convert it to a white image with a black character. After that Tesseract OCR engine applies LSTM multilingual classifier to recognize each character and finally gives the text of the image or document. The Figure 8 presents the mechanism of OCR.

To complete the tesseract operations, we have used the tesseract library, but tesseract does not perform for all images correctly. To get higher accuracy the images need to be pre-processed by rescaling, binarization, noise removing to avoid dropping tesseract output accuracy of letters, and digits. So, the system takes the image first then apply pre-processes algorithms such as for better accuracy image should be rescaled, converts image into grayscale, then apply dilation and erosion technique to remove noise from the image, binarize image using OTSU's algorithm. In our proposed system we have used the English and Bengali dataset to recognize characters.
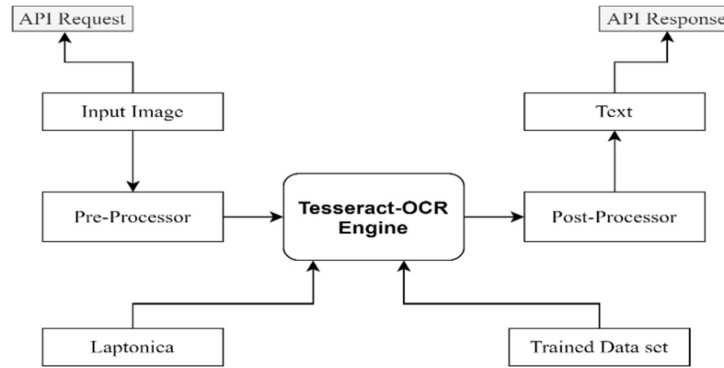
Figure 8. Diagram of the process of how tesseract engine detects the characters

2.5. Speech processing

Speech processing is introduced to control and communicate with the system. Formal English language is necessary to give a command to machine. It uses speech recognition techniques sometimes it's called automatic speech recognition (ASR); it is a subfield of computer science and computational linguistics that develop techniques that help to recognize and translate spoken language into text by machines or we can say it is a speech-to-text, enables a program to process human speech into written format.

The system takes voice commands from the person and transfers them into machine-readable that can execute the corresponding action. These commands, such as where he is, what objects in front of him, measure distance, which currency is, and query of him. Taking this query and execution then return the results to him through voice command that the blind can understand. Figure 9 shows the speech recognition architecture that used in this article.
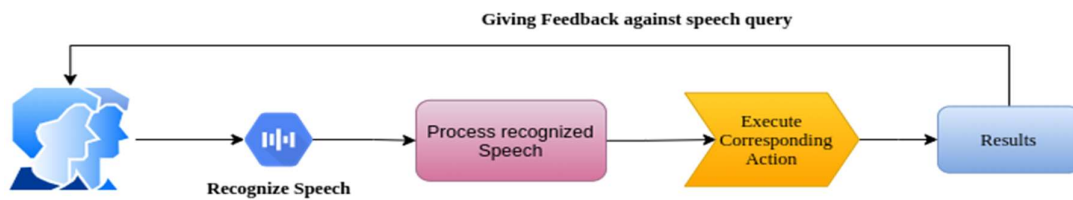
Figure 9. Working procedure of speech recognition module

We have created functions to respond user's query that means when the user speaks to do something then specific methods will be executed. Table 1 shows the common executable methods of the speech recognition module. There are also more methods to control the system and guide the blind person. Besides the above integrations, this system uses a GPS module that can locate blinds. A web application and android application is developed by providing adequate safety of blind's traveling and guide can monitor him from anywhere. Using these applications, the staff or guide can communicate with the blind when he or she faces complex or emergency situations.

Table 1. The functionalities of the speech recognition module

Speech	Actions	Output
'Start Device'	'The system starts'	System has started
'Where I am'	'getLocation () executes'	Current location
'Start Indoor Navigation'	'indoorNavigation () executes'	Started indoor navigation
'What is it?'	'findObjects ()'	Identify nearest objects
''Distance from objects'	'ultrasonicDistance ()'	Distance from the objects
'Where I am'	'The getLocation () method called	Current location's name

3. RESULTS AND DISCUSSION

To evaluate the overall system experimental results, several steps of the experiment have been driven such as perception tests, object detection, currency recognition, optical character recognition, navigation test in indoor and outdoor, GPS module, the next one is device control, including speech recognition and the last one is cloud environment, including web application and android application.

3.1. Object detection and recognition

The proposed system can identify 80 common objects, the detection and recognition results are shown in Figure 10. We have tested our detection module with different objects in both indoor and outdoor environments and got good results with higher accuracy. This module can also detect what the good exactly is and it performs well in both indoor and outdoor environments. The following figure also shows the confidence score to determine how accurate the objects are. There are many works has been done on object detection in [17]-[20]. We have shown some common object detection and recognition result.

Figure 10. Object identification with confidence score at (a) indoor and (b) outdoor

3.2. Currency recognition

To detect and recognize Bangladeshi currency we have developed a currency recognition module. It can recognize all Bangladeshi currency including coins and notes. The recognition results are shown in Figure 10 and it provides higher accuracy of identification. However, this currency recognition model cannot identify where the detected currency is fake or not. In the future, we will use ultraviolet technology to identify the authenticity of currency and also include other country's currencies. Figure 11 shows the recognition notes with the accuracy level.
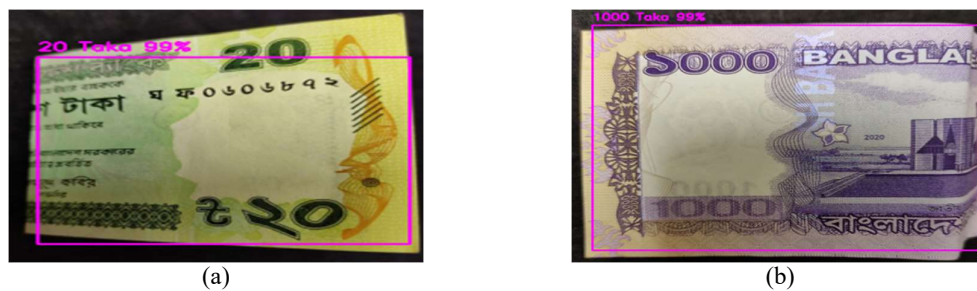

Figure 11. Currency recognition and the accuracy for (a) 20 taka (BDT) and (b) 1000 taka (BDT)

3.3. Face recognition

The face recognition module can detect known faces that are previously trained by the system. The accuracy of detection is good. The face detection and recognition are shown in the following figure. Face detection can be told as relative identification or friend identification. If any known faces come in front of the blind the system can identify and can tell who they are. Figure 12 represents that the system can identify the known faces.

Figure 12. The system can identify the known faces

3.4. Distance measure

Distance measurement can be divided into two parts, one is measuring the distance from the nearest object using an ultrasonic sensor and another one is measuring the distance from the person if any person's face is available in front of a blind camera. The system can perform calculating distance using the focal length of the object. Ultrasonic gives the distance from the object and sensor that is shown in Figure 13(a) and the distance using the camera from the person that is shown in Figure 13(b).
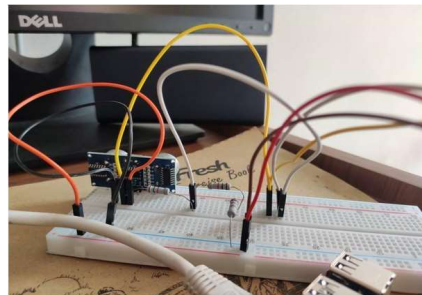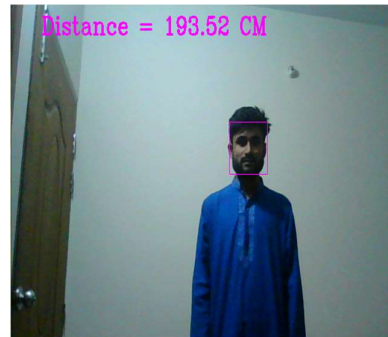
(a)   (b)

Figure 13. Distance measurement using (a) ultrasonic distance sensor from the wall and (b) person distance from the camera using CMOS

3.5. Optical character recognition

Recent developed OCR are basically twofold: one is text recognition in natural scene images and second one is document OCR that deals with document images, interested reader can refer [21]-[24] more elaborate discussion on these technologies. In our paper, we have implemented text recognition in document OCR.

To recognize English, Bengali, and digital numbers we have implemented ocr or we can say document reader whether the document is typed, scanned or photo that contains document. As shown in Figure 14, even if the images are blurred, the recognition results are still very accurate. Figure 14 shows the module can detect two different languages (a) for English and (b) for the Bengali language. It also gives good results on different types of images like jpg, png, jpeg, and it works on pdf also for extracting data.

(a)   (b)

Figure 14. The OCR document reader for (a) English language and (b) Bengali language

3.6. Speech processing

To control and interact with the system we have developed a speech processing unit that might give blind more feasibility. The module works with the natural English language. The specific function is being called with respect to a specific command. In the future, we will implement the recognition of the voices of friends and relatives. In Figure 15 we show the result of the query which was said by the user.

Figure 15. Speech processing and query result

3.7. Visualization of real-time application

The cloud is a way to manage and store data for blind interaction. We have developed a web application to track blind people in real-time. The guide can track their blind person; they can see the history of the visited area. They can also contact the help for any update for their device. Figure 14 shows the web application and android application interfaces. In both cases, the system can send the location data into the cloud using a GPS module that is attached with Raspberry pi. While visiting the guide can track his location by the web application (Figure 16(a)) and using the Android application they can also be able to track where the blind was visited in which area (Figure 16(b)). For doing booth application we have used Google API to implement these features.
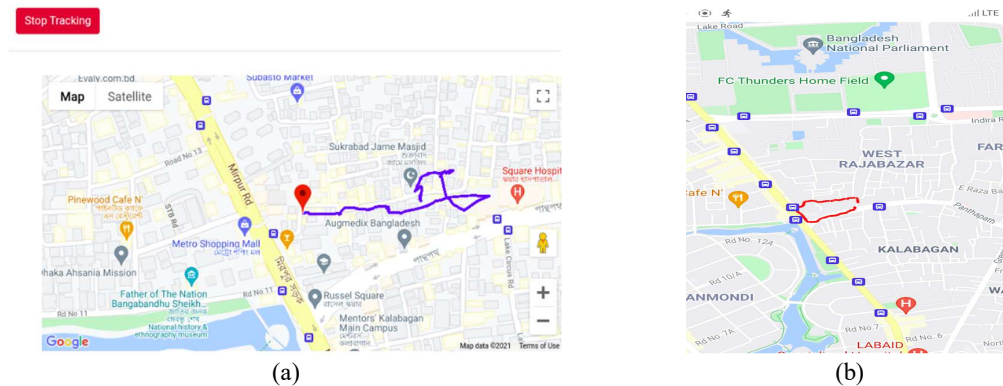

Figure 16. Real-time tracking overview of (a) web application and (b) android application

3.7. Work analysis with existing developed system

From the work of other researchers, we can confidently say that our work is absolutely the best of the others. In our proposed system, we were working with a huge amount of data with different types of identification. We worked on object detection, voice recognition, relative identity, human distance/focal distance, and so on. And the accuracy rate of it is also very highly good. In other papers, they just work with a part of it but we work for all the components. Table 2 shows the comparison of our work to other existing work.

Table 2. The comparative analysis of existing works

Author	Object detection	Speech	Face recognition	Person distance	Accuracy
Proposed system	✓	✓	✓	✓	98.9%
Duani *et.al*. [3]	✓	✓	✓	✗	-
Bai *et al*. [4]	✓	✓	✓	✗	99%
Bornschein *et al*. [5]	✓	✓	✓	✗	-
Sivan *et al*. [6]	✓	✓	✓	✗	97.5%
Setiadi *et al*. [7]	✓	✓	✓	✗	-
Yánez *et al*. [25]	✓	✓	✓	✗	-
Anandan *et al*. [26]	✓	✓	✓	✗	-
Hasan *et al*. [27]	✓	✓	✓	✗	-

4. CONCLUSION

We have developed a real-time mobile application as well as a web application for navigating visually impaired people. We used the updated version of the hardware, sensors, software tools, and libraries. The proposed system provides object detection, distance measuring, face, and relative recognition, OCR, speech recognition, GPS module integration (tracking), API building, Consequently, it could assist the visually impaired person for navigating in an indoor environment and outdoor environment. Furthermore, it can ensure the safe navigation of blind people in some complex and emergencies with an emergency contact method. The achieved accuracy of our system is 98% which is promising enough to ensure the quality of the system compared to existing systems. Though this system will play a significant role there are still some fields that need to be improved. In the future, we will work on a broader range of sensors and functionalities to improve the efficiency of our developed system.

REFERENCES

[1]   World Health Organization, ''Global data on visual impairment,'' 2021. [Online]. Available at: https://www.who.int/blindness/publications/globaldata/en/. [Accessed 24 March 2021].

[2]   J. U. Ahmed, S. Ahmed, and M. A. Majid, ''Caring Vision: Fighting Against Blindness in Bangladesh,'' *FIIB Business Review*, vol. 8, no. 1, pp. 17-24, 2019, doi: 10.1177/2319714519831545.

[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*, 2012, pp. 4208-4215, doi: 10.1109/IECON.2012.6389214.

[4]   J. Bai, D. Liu, G. Su, and Z. Fu, "A cloud and vision-based navigation system used for blind people," *Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies*, no. 22, pp. 1-6, 2017, doi: 10.1145/3080845.3080867.

[5]   J. Bornschein, D. Bornschein, and G. Weber, "Blind Pictionary: Drawing Application for Blind Users," *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1-4, 2018, doi: 10.1145/3170427.3186487.

[6]   S. Sivan and G. Darsan, "Computer vision based assistive technology for blind and visually impaired people," *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies*, pp. 1-8, 2016, doi: 10.1145/2967878.2967923.

[7]   B. Setiadi, T. Supriyadi, H. Nugroho, and R. Solihin, "Navigation and Object Detection for Blind Persons Based on Neural Network," *International Journal Applied Technology Research*, vol. 1, no. 1, pp. 56-65, 2020, doi: 10.35313/ijatr.v1i1.24.

[8]   K. Sarakhman, R. Kempnyk, and V. Chyhura, "Object Detection and Descriptive Analytics," *Computational linguistics and intelligent systems: proceedings of the 4nd International conference*, vol. 2, pp. 425-428, 2020.

[9]   P. A. Majid, V. Nikhitha, K. M. Smrithi, and V. P. Basheer, "Machine Learning Based Mobility Aid for Blind People," *Machine Learning*, vol. 7, no. 1, 2020.

[10]   K. Vijlyakumar, K. Ajitha, A. Alexia, M. Hemalashmi and S. Madhumitha, "Object Detection for Visually Impaired People Using SSD Algorithm," *International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2020, pp. 1-7, doi: 10.1109/ICSCAN49426.2020.9262344.

[11]   R. Abirami, S. Haarini, and P. Hari, "Customized Smart Glasses for Needy Blind People," *The Learning and Technology Library*, pp. 33-44, 2020.

[12]   A. Ismail, S. Abdlerazek and I. El-Henawy, "Development of Smart Healthcare System Based on Speech Recognition Using Support Vector Machine and Dynamic Time Warping," *Sustainability*, vol. 12, no. 6, pp. 1-15, 2020, doi: 10.3390/su12062403.

[13]   A. Bansal, "Ultrasonic Distance Meter Using Raspberry Pi 2 | Electronics Project," *Electronics for You*, 2021. [Online]. Available: https://www.electronicsforu.com/electronics-projects/ultrasonic-distance-meter-raspberry-pi-2.

[14]   A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv [cs.CV], 2020.

[15]   "COCO-Common Objects in Context," *Cocodataset.org*, 2021. [Online]. Available: https://cocodataset.org.

[16]   A. Geitgey, "Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning," *Medium*, 2021. [Online]. Available: https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78.

[17]   M. T. Habib, M. J. Mia, M. S. Uddin, and F. Ahmed, "An in-depth exploration of automated jackfruit disease recognition," *Journal of King Saud University-Computer and Information Sciences in Press*, 2020, doi: 10.1016/j.jksuci.2020.04.018.

[18]   M. T. Habib, M. R. Mia, M. J. Mia, M. S. Uddin, and F. Ahmed, "A Computer Vision Approach for Jackfruit Disease Recognition," *Proceedings of International Joint Conference on Computational Intelligence*, Springer, Singapore, 2020, pp. 343-353, doi: 10.1007/978-981-15-3607-6_28.

[19] J. Mia, H. I. Bijoy, S. Uddin and D. M. Raza, "Real-Time Herb Leaves Localization and Classification Using YOLO," *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1-7, doi: 10.1109/ICCCNT51525.2021.9579718.
[20] Md. J. Mia, S. K. Maria, S. S. Taki, and Al A. Biswas, "Cucumber Disease Recognition Using Machine Learning and Transfer Learning," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3432-3433, 2021, doi: 10.11591/eei.v10i6.3096.
[21] Q. Ye and D. Doermann, "Text Detection and Recognition in Imagery: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 7, pp. 1480-1500, 2015, doi: 10.1109/TPAMI.2014.2366765.
[22] J. Ma *et al*., "Arbitrary-Oriented Scene Text Detection via Rotation Proposals," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3111-3122, 2018, doi: 10.1109/TMM.2018.2818020.
[23] T. M. Breuel, "High Performance Text Recognition Using a Hybrid Convolutional-LSTM Implementation," *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, pp. 11-16, doi: 10.1109/ICDAR.2017.12.
[24] M. Yousef, K. Hussain and U. Mohammed, "Accurate, data-efficient, unconstrained text recognition with convolutional neural networks," *Pattern Recognition*, vol. 108, p. 107482, 2020, doi: 10.1016/j.patcog.2020.107482.
[25] D. V. Yánez, D. Marcillo, H. Fernandes, J. Barroso, and A. Pereira, "Blind Guide: anytime, anywhere," *Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pp. 346-352, 2016, doi: 10.1145/3019943.3019993.
[26] M. Anandan, M. Manikandan, and T. Karthick, "Advanced Indoor and Outdoor Navigation System for Blind People Using Raspberry-Pi," *Journal of Internet Technology*, vol. 21, no. 1, pp. 183-195, 2020.
[27] Md. Hasan, F. Ahmmed, and Md. Haque, "A Method for Object Detection to Help Blind People Using Machine Learning," *DSpace Repository*, 2021.

BIOGRAPHIES OF AUTHORS

Md. Atiqur Rahman received his Engineer degree (B.Sc.) in Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh in 2021. Currently, he works as a Software Engineer at ThemeXpert Ltd, Dhaka, Bangladesh. He worked at Nhimex Fuel Managing Company, Dhaka, Bangladesh, as Jr. Software Engineer and worked on IoT based research and development wing and he also worked on design, development of web application. His research interest is mainly on Artificial intelligence, computer vision, machine learning, IoT, data mining, and natural language processing. He can be contacted at email: atiq.misterrahman@gmail.com.

Sadia Siddika received her Engineer degree (B.Sc.) in Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh in 2021. Currently, she is working on web application development and her machine learning research will be applied on that. Her research interest is mainly on computer vision, machine learning, data mining. She can be contacted at email: sadiasiddika9913@gmail.com.

Md. Abdullah Al-Baky received his Engineer degree (B.Sc.) in Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh in 2021. Currently, he works as an Android Developer at Karzy IT, Dhaka, Bangladesh. His research interest is mainly on artificial intelligence, computer vision, machine learning. He can be contacted at email: abddiucse@gmail.com.

Md. Jueal Mia He received his B.Sc. (Honors) and M.Sc. in Computer Science and Engineering from Jahangirnagar University, Dhaka, Bangladesh in 2014 and 2015 respectively. Currently, he works as a Senior Lecturer in the Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. He has published many articles in international journals and conference proceedings. His research interest is mainly on artificial intelligence, computer vision, machine learning, expert systems, data mining, and natural language processing. He can be contacted at email: mjueal02@gmail.com.