

Received December 29, 2021, accepted January 24, 2022, date of publication January 31, 2022, date of current version February 10, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3148036

# CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People

FAHAD ASHIQ<sup>1</sup>, MUHAMMAD ASIF<sup>1</sup>, MAAZ BIN AHMAD<sup>2</sup>, SADIA ZAFAR<sup>1</sup>,  
KHALID MASOOD<sup>1</sup>, TOQEER MAHMOOD<sup>3</sup>, MUHAMMAD TARIQ MAHMOOD<sup>4</sup>, (Senior Member, IEEE),  
AND IK HYUN LEE<sup>5</sup>

<sup>1</sup>Department of Computer Science, Lahore Garrison University, Lahore 54000, Pakistan

<sup>2</sup>College of Computing and Information Sciences, Karachi Institute of Economics and Technology, Karachi 75190, Pakistan

<sup>3</sup>Department of Computer Science, National Textile University, Faisalabad 37610, Pakistan

<sup>4</sup>Future Convergence Engineering, Korea University of Technology and Education, Byeongcheonmyeon, Cheonan 31253, Republic of Korea

<sup>5</sup>Department of Mechatronics Engineering, Korea Polytechnic University, Siheung, Gyeonggi-do 15073, Republic of Korea

Corresponding author: Ik Hyun Lee (ihlee@kpu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant by the Korean Government through the Ministry of Science and ICT (MSIT) under Grant 2021R1F1A1050022, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) by the Ministry of Education under Grant 2017R1A6A1A03015562.

**ABSTRACT** Visually impaired persons (VIPs) comprise a significant portion of the population, and they are present around the globe and in every part of the world. In recent times, technology proved its presence in every domain, and innovative devices assist humans in their daily lives. In this work, a smart and intelligent system is designed for VIPs to assist mobility and ensure their safety. The proposed system provides navigation in real-time using an automated voice. Though VIPs wouldn't be able to see objects in their surroundings, they can sense and visualize the roaming environment. Moreover, a web-based application is developed to ensure their safety. The user of this application can turn the on-demand function for sharing his/her location with the family while compromising privacy. Through this application, the family members of VIPs would be able to track their movement (get location and snapshots) while being at their homes. Hence, the device allows VIPs to visualize the environment and ensure their security. Such a comprehensive device was a missing link in the existing literature. The application uses MobileNet architecture due to its low computational complexity to run on low-power end devices. To assess the efficacy of the proposed system, six pilot studies have been performed that reflected satisfactory results. For object detection and recognition, a deep Convolution Neural Network (CNN) model is employed with an accuracy of 83.3%, whereas the dataset contains more than 1000 categories. Moreover, a score-based quantitative comparative analysis is performed using the supported features of devices. It is found that the proposed system has outperformed the existing devices having a total score of 9.1/10, which is 8% higher than the second-best.

**INDEX TERMS** Object detection, security surveillance, visually impaired persons, CNN, deep learning.

## I. INTRODUCTION

The World Health Organization (WHO) has reported that 285 million of the world's population is blind or visually impaired. Out of these, 39 million are blind [1]. The major diseases that cause visual impairments include refractive error, glaucoma, trachoma, corneal opacities, cataracts, diabetic retinopathy, and unaddressed presbyopia [2].

Visually impaired persons (VIPs) face difficulties in performing activities of daily living (ADLs) e.g. opportunity of work and schooling, moving in their surroundings, ability

The associate editor coordinating the review of this manuscript and approving it for publication was Oğuzhan Urhan<sup>1</sup>.

to interact with the environment, and searching for common objects (indoor/outdoor) at their own or even with some assistance. The main challenges for VIPs are/object detection and recognition, currency identification, textual information (sign, symbol) and translation, mobility/navigation and safety [3].

In the past, several approaches, systems, devices, and applications have been developed in the domain of assistive technology to facilitate VIPs to perform tasks that they were formerly unable to accomplish [4]. Such solutions generally comprise electronic devices equipped with cameras, sensors, and microprocessors capable of making decisions and providing tactile or auditory feedback to the user. Many of

the existing object detection and recognition systems claim high accuracy but cannot provide necessary information and attributes for the tracking of VIPs to ensure their safe mobility [5].

Though blind people cannot see objects in their surroundings, it would be helpful to know about them. Furthermore, there is a need to develop a tracking system through which family members of VIPs can monitor their movement.

In the light of the requirements above, this paper presents a smart system that performs real-time object localization and recognition. As soon as the system recognizes the object, it sends audio feedback to the user. For example, after identifying a known object (e.g., car), the user will hear the word “car.” Moreover, the user’s location and a snapshot of the most recent viewed scene are periodically stored at a server that can be accessed by the family members using an application to track the user. For object detection and recognition, mobileNet architecture is used because of its low computational complexity to run on low power end devices. Since wearable hardware resources are limited, and the system’s feedback about the object’s name needs to be as close as possible, complex state-of-the-art object recognition approaches might not be viable as the first choice techniques.

The main objective of this study is to develop a system for VIPs that contains the following features:

- It performs real-time object detection and recognition using a deep learning framework.
- It pronounces names of the objects that can be seen through the camera eye i.e., objects present in the current frame.
- It periodically sends the location of the user to a web server.
- It sends live feed/snapshots to the webserver.
- It provides a web-based interface that the family members of VIPs can use to track their movement while being at their homes. It is an on-demand feature that provides security and ensures the user’s privacy.

The rest of the article is organized as follows: Literature review is performed in section 2. Section 3 presents the proposed framework. The experimental analysis is discussed in section 4 and, the conclusion is drawn in section 5.

## II. LITERATURE REVIEW

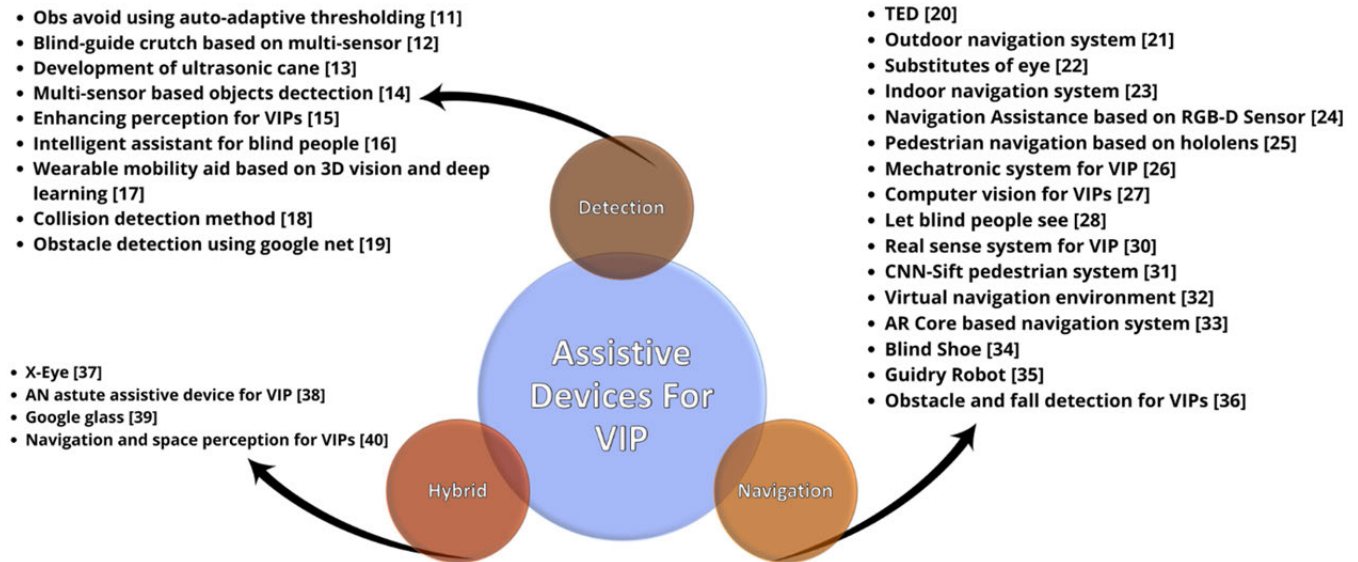
The researchers have several proposed techniques to develop assistive devices for VIPs. The following technologies have been used in the existing devices: vision-based sensors (camera), non-vision-based sensors (e.g., IR, ultrasonic, inertial and magnetic sensing, etc.), and other technologies such as low-energy Bluetooth beacons, GPS, GPRS, etc. In this section, the main focus is put on vision-based sensing devices due to their relevance to our proposed system. The existing assistive devices can be grouped into the following three classes according to their functionalities:

- Object detection devices
- Navigation devices

- Hybrid devices (Object detection and navigation)
- The devices can be sub-classified based on their working principle and by means (sound or vibrations) of conveying information to VIPs. Figure 1 presents the taxonomy of the assistive devices for VIPs.

The object detection devices use sensors (laser scanners, ultrasonic devices) and cameras to collect information from the surrounding environment, process it, and provide feedback to the users. The basic working principle of such devices is that they detect the object around the user and give instructions about the object/obstacle and its distance using vibrations or sound waves. Saputra *et al.* [6] have presented an obstacle avoidance system with the help of a Kinect depth camera for VIPs. It helps detect the obstacle and calculate its distance using the auto-adaptive threshold. The device is tested on ten blind people aged 20-40 years to evaluate the system’s performance. The result of the proposed system is promising as it detects the obstacle without any collision from any direction. Yi and Dong [7] have presented a blind-guide crutch using multiple sensors. The triplet ultrasonic module detects the obstacle from the front, left, and right sides. It identifies the object using the voice and vibration waves. Kumar *et al.* [8] have presented an ultrasonic cane that provides information related to the environment and enables the user to move safely. The ultra-Cane consists of a narrow beam ultrasound system that provides 100% obstacle detection. It detects objects which are 2- 4 meters away. The proposed device is tested on ten people (ages 20-26 years). The volunteers effectively detected the hurdles within the proposed range. Petal *et al.* [9] developed a multi-sensor system that helps the user detect an object in the indoor surroundings. The system uses statistical parameters and the SVM (Support Vector Machine) algorithm for detecting the object. The response is given using the audio. Bauer *et al.* [10] have introduced a device based on the wireless cam for capturing the atmosphere using the depth map for surroundings. The object detector detects the semantic of an object, whereas the depth map provides the 3D information of the surrounding. The haptic or spoken feedback is given to the user.

The application introduced by Chen *et al.* [11] is helpful in the detection of obstacles and contains glasses, a long stick cane, and a mobile application. The object can be detected through glasses and the long stick cane. If the user falls, the information is uploaded on the online platform and is displayed on the mobile device. Pogi and Mattocchia [12] have developed a wearable system for VIPs using deep learning and 3D vision based on the smartphone. The system uses an RGB camera to capture frames and CNN to detect obstacles. The system’s performance is exceptional, i.e., near to 98% through the LeNet architecture. The system is lightweight, small in size, and performs detection in real-time, but it cannot categorize most of the daily life objects. The user is guided through the tactile and audio feedback. The system is tested on more than 20 individuals having a visual impairment of different degrees. The result is very



**FIGURE 1.** Taxonomy of assistive devices for VIPs.

promising with good detection performance. Chae *et al.* [13] developed a system for detecting the collision through image segmentation. It detects obstacles that the existing solutions cannot be detected i.e., walls, doors etc. The input frames are segmented into superpixels through the lattices algorithm. For segmentation results, a Graph-based merging algorithm is used to detect both non textured and textured objects successfully. The system is tested on sighted/blindfolded people, dynamic/static objects. The experiments result exhibit that the proposed system efficiently detects the object in real-time.

Salavati and Mahvash [14] have introduced a system for detecting obstacles through a camera based on the Deep Neural Network (DNN). The system comprises unsupervised deep neural networks for extracting global features of the image, whereas, for extracting local attributes, a supervised DNN is used. Nguyen *et al.* [15] have proposed an electro tactile device connected with the tongue that helps the user navigate. The device is supported by an antenna that enhances wireless communication. Xiao *et al.* [16] have introduced an outdoor assistive navigation system controlled by the Raspberry Pi and GPS sensor, guiding navigation. The device has a list of the recordings. The user selects any sound from the list to receive the navigation. The device also has a speech recognition system to enhance its capabilities. The system is tested in New York City College to evaluate its performance. The result shows that every direction is followed properly using the developed device. Bharambe *et al.* [17] have presented a device that uses an ultrasonic sensor for navigation and direction. This device is based on GSM and GPRS. GPS is used to receive the specific location of a particular person. The device is tested on blindfolded persons to evaluate their performance. Martinez-Sala *et al.* [18] have developed a sugar navigation device for VIPs. The designed device is tested on blindfolded persons. The system uses ultra-wide techniques that provide accurate information for the location. Its accuracy is about 95%. Aladren *et al.* [19] have introduced a

system based on the RGB-D sensor that guides navigation. The RGB-D is used for obtaining the range and the color information. The individual wears the device to evaluate the performance. The results show that the algorithm is robust and gives approximately 95% accuracy. Yamashita *et al.* [20] have introduced a wearable device that helps VIPs navigate using RFID, Quasi-Zenith system, and Hololens of Microsoft. The Hololens contain multiple cameras and sensors which help in positioning. Four experiments have been conducted against each individual to assess the system's performance. So, in total, 16 versions of different scenarios were attained. Out of these scenarios, users could move on the right path without any collision in thirteen situations. Mancini *et al.* [21] have offered a vision-based device that guides VIP in jogging, walking, and running. The system consists of gloves, a camera, and a board equipped with motors. The system recognizes the accurate path with speeds around 10 km/h using gloves.

Caraiman *et al.* [22] have proposed an accurate and robust approach for segmentation and reconstruction of the outdoor surroundings using IMU and SoV camera data. It helps in the mobility of a user in the outdoor environment. The system provided a 3D environment even though the data was noisy. The system's performance is demonstrated through a large dataset recorded using a camera. The information is conveyed to the user through an audio and haptic interface. "Let Blind People See", a system proposed by Jiang *et al.* [23] to detect obstacles, and audio feedback is used to convey an alert to the users. The mobile camera has been used to record the video used in the input for you only look once (YOLO) [24] algorithm for detecting the obstacles in real-time. Moreover, the engine of a 3D game-based application sends and creates 3D locations and audio for objects. The device is tested on a normal user whose eyes are covered with paper to evaluate the system's performance. The user successfully identifies the objects. Yang *et al.* [25] presented a wearable device that

helps detect the traversable area through RGD-Camera based on the random sample consensus (RANSAC) and surface normal vector estimation. The device gives the interface of non-semantic for transmitting the results to VIPs. It works in both indoor and outdoor surroundings. However, due to heavy computation, it's difficult for a device to work in a real-time environment and it detects only a short range of attainable areas. For assessing the performance of the system, it is tested on eight individuals and three of them are blind. The results display that the system is useful in an unknown environment.

Zhao *et al.* [26] introduced a hybrid system utilizing CNN and features of an image to get first-person vision (FPV) navigation. The system takes a sequence of videos as input and gives output in a map showing the user's position. The process of this system is divided into two sub-processes, i.e., estimation of a location and calculation of a movement. An AlphaMex-based on CNN is developed to recognize scenes in real-time, and SIFT tracking algorithms are proposed to analyze the movement. The IMU sensor easily influences the noise, and the complete navigation system depends on the single camera instead of using GPS or any other sensors. Kunz *et al.* [27] have introduced the idea of Virtual Navigation to aid VIPs in facing problems in mysterious locations through a real walk while staying in a precise environment. For achieving this, a user needs a camera on the head and brings the laptop to the back. Zhang *et al.* [28] developed a navigation system for VIPs using ARCore and provided safety and efficiency. The dual-channel mechanism that interacts with humans is introduced to deliver continuous and accurate direction that ensures path safety from risky areas and obstacle detection. The information is conveyed to the user through an audio and haptic interface.

For evaluation of the system, different experiments have been conducted. The results show that the proposed system is much more efficient than the already available system. It tested on one blind and three vision impaired subjects to assess the system's performance. After the test, a survey was conducted, and the result shows that this entire subject efficiently follows the path to reach its destination. Kubanek and Bobulski [29] have developed an electronic device that helps in the orientation of VIPs in the surrounding through voice signals. The device comprises a memory card, helmet, plug, cables, sensor, raspberry pi, headphones, and resistor for the Kinect sensor. The area in front of the user is scanned through the Kinect sensor, and also different algorithms are used for generating the acoustic space. The experiments have been performed on people without visual impairment to check the efficiency of the system. The results displayed that the performance of the presented system is high. Rahman *et al.* [30] have developed an Electronic Assistance system for VIPs that helps them navigate. It comprises the data transmitter device, sensors, micro-controllers, and smartphones. The system is associated with the smartphone. The information is collected through sensors from the environment; the obstacle is detected through a microcontroller after processing the collected information.

Bluetooth module helps in communication between smartphones and micro-controller. The smartphone provides the instructions continuously to keep the user on track. The device is tested in a real-time environment for evaluating performance. The obstacle detection is performed at 100cm with 97.33% accuracy.

Megalingam *et al.* [31] have designed a robot that acts as path guidance for blind persons. It is an addition to the guide dog. The robot can move on multiple paths. It can also locate those places that are not even traced by the GPS module. The proposed system comprised the map, camera, and image processing algorithms. The robot helps the users in tracking their location. The robot is operated through a joystick that converts axes into voltage, and the ultrasonic sensor is used to find the distance in the range of 2-3 cm. It is placed on the motor for detecting the obstacles. The IR sensor is used to help in measuring the depth. The response is conveyed to the user through vibration. It is helpful for both indoor and outdoor environments. The experiments are performed on real-time data that give accurate results. Rahman *et al.* [32] have proposed a wearable system that helps VIPs to move independently. The system helps detect an obstacle that is placed in front of the user. The system consists of a motion sensor, smartphone application, accelerometer, ultrasonic sensor, microcontroller, and the device for data transmission. The audible instruction is produced to assist the user in navigation. The data is transmitted through a microcontroller to the smartphone using Bluetooth. It also helps in updating the current location and keeps them on track. In case, if the user falls, it notifies the guardians. The performance is evaluated by testing it on the target users. The result shows an accuracy of 98.3% for the proposed system.

Minhas and Javed [33] have developed X-Eye, which facilitates detection and recognition of an object and helps trace a location. It is composed of a camera and a smartphone. Meshram *et al.* [34] have introduced NavCane, which helps the VIPs based on the communication module, ultrasonic sensor, and the motor for vibration. It helps in guiding the accurate path and also in the detection of an object. It works only in indoor surroundings. To evaluate the system's performance, 80 users participated in the experiments. Out of the total number, 30 participants were blind, 30 had low-vision, and 20 were older adults. The result shows that the proposed system helps in navigation. Berge *et al.* [35] have introduced a device that helps VIPs and other disabled people through an Android-based system. It is helpful not only for detecting obstacles but also provides navigation.

The system was tested on 15 users. 10 of these were experiencing disabilities of visual impairment, and 5 of them were blind. All of these participants tested the system in indoor/outdoor surroundings. The results of the proposed system are very promising as it successfully helps in navigation. Kammoun *et al.* [36] have invented a system called NAVIG to aid VIPs. It contains microphones, cameras, headphones, GPS, sensors, and a computer. The device uses a stereoscopic camera located at the top side to capture the images from the



environment. Machine learning techniques are used for object detection. The NAVIG works both in indoor and outdoor surroundings. The results demonstrated that the system is working correctly.

Chang *et al.* [38] introduced a wearable system that assists VIPs using artificial intelligence (AI) techniques. The system guides VIPs in zebra crossing and marked crosswalks. The system comprises a waist-mounted device, a walking cane, and sunglasses. A deep learning method is used to recognize zebra crossing in real-time. When VIPs reach a zebra crossing, they will instantly receive a message whether to move or stop. The system achieved accuracy up to 90%. Chang *et al.* [38] developed an intelligent obstacles detection system for VIPs. The system consists of a walking stick, smart glasses, a cloud-dependent platform, and a mobile application. VIPs wear smart glasses and hold the walking stick to identify obstacles. The developed stick starts vibrations for the guidance of VIPs. Furthermore, a notification is sent to their registered family member in case of any accidental event. The authors claimed that the system is 98.3% accurate. Rahman and Sadi [39] developed a system for VIPs that helped them identify outdoor and indoor objects through audio messages. The system is based on four sensors that help detect the object from all directions. The overall accuracy of the system is approximately 99%.

In previous years, numerous studies have been performed to address the challenges, i.e., understanding the environment, increasing perception, the nature and position of an obstacle on the travel path, and facilitating VIPs in outdoor and indoor environments. Mostly, the developed technologies include functions for selecting routes and detecting an obstacle [41].

But still, there is a need to design and develop intelligent systems that help in object detection and recognition and facilitate securing the mobility of VIPs.

### III. THE PROPOSED FRAMEWORK

The proposed system comprises a raspberry pi digital signal processing (DSP) board with GSM and a global positioning system (GPS) module, headphones, and a camera. DSP captures a live feed from the video camera and passes it to the object detection and recognition module (CNN model). The model predicts objects in the current video frame and passes their names to the text to speech converter (SAPI) module, pronouncing their names using headphones. Moreover, the labeled snapshot is encoded using the Joint Photographic Experts Group (JPEG) encoder. The encoded image with the user's precise location is saved in the server's database. Furthermore, a user-driven feature is provided to VIPs that enables family members to track their movement (get current labeled frame and location) while relaxing at their home through a web interface. The user can enable/disable this feature through a single button press. Figure 2 provides the system architecture of the proposed solution that comprises the following main modules:

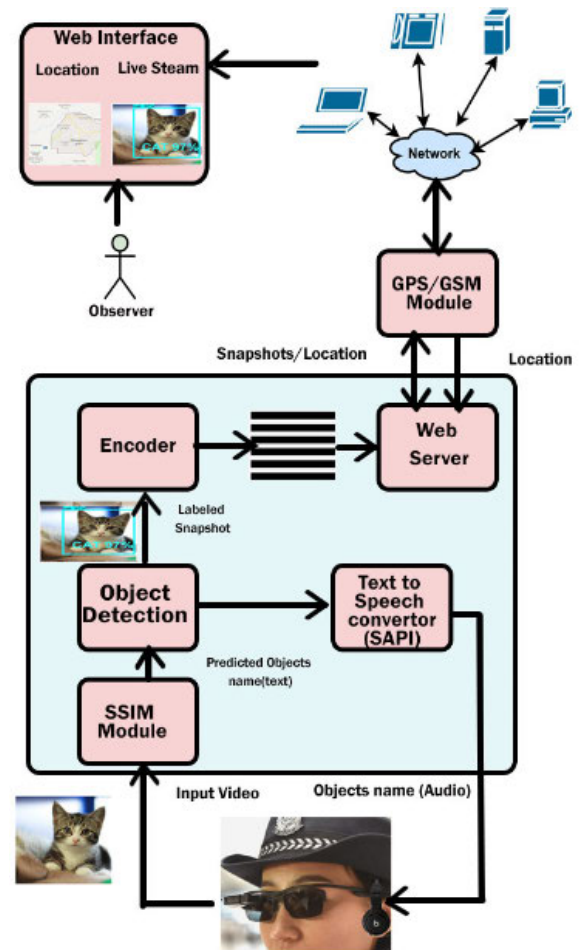


FIGURE 2. The architecture of the proposed system.

#### A. DSPBoard (Pi 3 MODEL B)

The Raspberry Pi foundation has developed a number of small on board computers to handle limited control functions. A system on a chip (SoC) along with integrated Acom RISC machine (ARM) and a robust central processing unit ICPU are installed on Raspberry Pi. We have used raspberry Pi 3 Model for the development of the proposed system [42]. The size of the device is similar to the standard credit card and it is the third generation of the series. Its primary specifications contain ARM Cortex A53 (Broadcom BCM2387) quad core processor with 64 bit CPU at 1.2 GHz speed and 1 GB of RAM. There are 4 poles stereo output and CSI camera port along with composite video port that is used to connect a full HDMI camera that is accessible to wireless LAN (802.11 b/g/n) and Bluetooth 4.1 chipset.

#### B. SOFTWARE BASED OBJECT DETECTION AND RECOGNITION MODULE (SODRM)

The system employs Mobile-Net architecture [41] for object detection and recognition because it is light weighted and provides good accuracy that makes it the best choice for computationally low devices/DSPs. A convolutional neural network (CNN) is trained for object detection in a python programming language using tensor-flow API.

TABLE 1. The mobile net architecture.

Type/Strid	Filter Shape	Input Size
Conv / s2	3×3×3×32	224×224×3
Conv dw / s	3×3×3×32 dw	112×112×32
Conv / s1	1×1×32×64	112×112×32
Conv dw / s2	3×3×3×64 dw	112×112×64
Conv / s1	1×1×64×128	56×56×64
Conv dw / s1	3×3×128 dw	56×56×128
Conv / s1	1×1×128×128	56×56×128
Conv dw / s2	3×3×128 dw	56×56×128
Conv / s1	1×1×128×25	28×28×12
Conv dw / s	3×3×256 dw	28×28×256
Conv / s1	1×1×256×256	28×28×256
Conv dw / s2	3×3×256 dw	28×28×256
Conv / s	1×1×256×51	14×14×256
5 × Conv dw / s1	3×3×512 dw	14×14×512
Conv / s1	1×1×512×512	14×14×512
Conv dw / s2	3×3×512 dw	14×14×512
Conv / s1	1×1×512×1024	7×7×512
Conv dw / s2	3×3×1024 dw	7×7×1024
Conv / s1	1×1×1024×1024	7×7×1024
Avg Pool / s	Pool 7 × 7	7×7×1024
FC / s1	1024 × 1000	1×1×1024
Softmax / s1	Classifier	1×1×1000

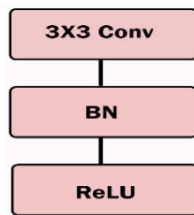


FIGURE 3. The architecture of simple convolutional layers.

C. MOBILE-NET ARCHITECTURE

Mobile-Net [41] is based on streamlined architecture to build a lightweight deep neural network. It consists of 21 layers. It uses depth-wise separable convolution layers that can be either simple or deep convolutional layers. Its architecture is defined in Table 1.

Mobile Net architecture consists of 21 layers containing both simple and deep convolutional layers. The architecture of simple convolutional layers is shown in Figure 3. Simple convolution layers consist of three steps. At each layer, convolution is applied using the 3 × 3 kernel followed by the batch normalization (BN) and rectified linear unit (ReLU) activation function.

On the other hand, the deep layers consist of six stages. The first three steps of deep and simple layers are the same. These steps are followed by the 1 × 1 convolutional operation, batch normalization (BN), and rectified linear unit (ReLU) activation function. The architecture of deep convolutional layers is shown in figure 4.

D. STRUCTURAL SIMILARITY (SSIM) INDEX

As in the video signal, there is a strong correlation among the frames belonging to the same scene. In this case, detecting,

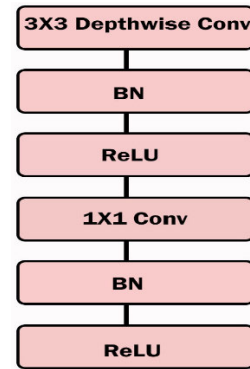


FIGURE 4. The architecture of deep convolutional layers.



FIGURE 5. GPS and GSM module chip.

recognizing, and pronouncing the same objects are not useful. It is better to check the similarity between two successive frames in the displaying order. If the similarity index is less than a certain threshold T, process the current frame. Otherwise, skip this frame (i.e., there is no need to process it). This skipping saves computational power and helps to develop a real-time system. The SSIM index [42] is used to measure the similarity between two successive frames in this work. It is used to decide whether the current frame is needed or not. The following expression measures the SSIM index between frames I and J of dimension M × N.

$$SSIM_{I,J} = \frac{(2\mu_I\mu_J + S_1)(2\sigma_{I,J} + S_2)}{(\mu_I^2 + \mu_J^2 + S_1)(\sigma_I^2 + \sigma_J^2 + S_2)} \quad (1)$$

where,

- $\mu_I$  and  $\mu_J$  represent the mean of the frame I and J, respectively.
- $\sigma_I^2$  and  $\sigma_J^2$  indicate the variance of the frame I and J, respectively.
- $\sigma_{I,J}$  is the covariance of the frame I and J.
- $S_1 = (C_1L)^2$  and  $S_2 = (C_2L)^2$  stabilize the division with a denominator. L indicates the dynamic range of pixel-values which is 255,  $C_1 = 0.01$  and  $C_2 = 0.03$ .

After detailed simulations on various video streams, the SSIM index threshold T is adjusted to 0.7. If  $SSIM > T$ , skip the frame; otherwise, process the frame.

E. TEXT TO SPEECH CONVERTER (SAPI)

In this work, the Speech API (SAPI 5.3) [43] is used to generate audio based on the input text. To pronounce the

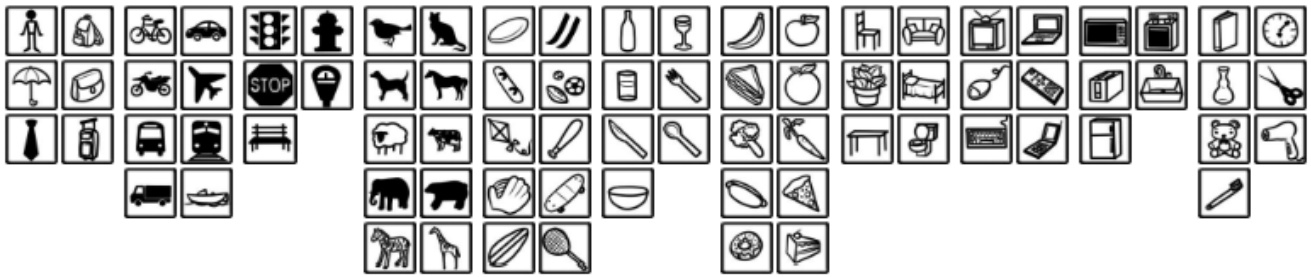


FIGURE 6. Sample objects in ImageNet dataset.

detected objects in the current frame, SAPI takes the textual input (consisting of objects names present in the current frame) from CNN and generates an audio signal. The output of SAPI is sent to headphones through which VIP scans hear the names of the detected objects.

#### F. ENCODER

The CNN provides labeled snapshots in which bounding boxes are drawn around the objects and labels/names are assigned. These images are encoded using the Joint Photo Experts Group (JPEG) encoder to store them in a database and access them through the web interface.

#### G. GPS AND GSM MODULE

The Pi Anywhere 4G and LTE Hat for the Raspberry Pi Beta are used in the proposed system [44]. The Raspberry Pi minicomputer provides 4G mobile data, GPS positioning information, and battery support. It is a simple plug and plays module. It provides ultra-fast 4G (100 Mbps down/ 50 Mbps up) Internet connectivity for video streaming and downloads. Moreover, it provides easy access to GPS onboard, exposing location data. The prime purpose of the module is to provide connectivity and the location of the user for tracking. Figure 5 shows the GPS and GSM module chips.

#### H. WEB-BASED APPLICATION

The motive behind the development of the web interface is to ensure the safety of VIPs. This interface facilitates the family members of the VIPs to track their movement (get location and snapshots) while sitting at home. The device allows VIPs to visualize the environment and ensure their safety. The web server is running on DSP. It periodically (in real-time) sends the person's position (GPS coordinates) and labeled snapshot to the web interface. The web interface maps the received coordinates on the map using Google-map API that provides the user's real-time location. Moreover, it displays the most recent received labeled image and stores all the received images in the gallery. The gallery is helpful to review the visited locality.

In this work, Django [45] is used to develop a web-based application. To monitor visually impaired persons, a user needs to login on to a web page via a web browser using the given ID. After signing in, the user is redirected to the dashboard, which contains the most recent snapshot received from the web-server and a real-time location of the VIPs. The web

interface will receive the live video feed from the camera and the person's coordinates in real-time. The received coordinates would be mapped on the map using Google map API that provides the real-time location of the blind person.

#### IV. TECHNICAL EVALUATION

This section presents scenario-driven evaluations of the prototype to test the efficacy of the proposed prototype. The idea is that blind people are concerned about objects present in their close vicinity. The system offers the presence of an object and conveys the type of an object through voice, and periodically sends the person's location to the webserver.

For object detection and recognition, Mobile-net is used because it is fast and requires less computational power. The model is trained on the ImageNet dataset that is large-scale object detection, segmentation, and a captioning dataset comprising 2.1 million images in 1000 categories [46]. The sample objects of the ImageNet dataset are shown in figure 6.

The training of Mobile-Net is performed on Ubuntu 16.04 LTS platform with Xeon E3-1231 v3, 4 Cores @ 3.40GHz, GTX1060 [47]. It achieves an accuracy of 83.3%. Table 2 illustrates the comparison of MobileNet architecture with other models in terms of accuracy, number of model parameters, error rate, and complexity in terms of Mega Floating-point Operations Per Second (MFLOPs) [48]. It is reflected that MobileNet architecture is 32 times smaller in terms of hyper-parameters and 27 times less computationally intensive than VGG [41]. However, the accuracy of shuffle-net [48] is higher than mobile-net. The proposed system is flexible enough to accommodate any model, including shuffle-net. In the future implementation, shuffle-net may be incorporated instead of mobile-net.

The average loss graph of the trained model is shown in figure 7. It is evident from the figure that after 800k iterations, loss becomes the minimum which is about 1.39.

The prototype works as follows: it takes input from the camera, performs object detection and recognition, and pronounces the name of objects that can be seen through the camera eye. The notification is sent to the user through headphones. The system can perform multiple object detection and recognition in a snapshot. In the case of multiple objects in one snapshot, the objects are sorted in descending order according to their confidence level/probability of prediction. The object with a higher confidence level is pronounced first and subsequently the rest of the objects. In the case of



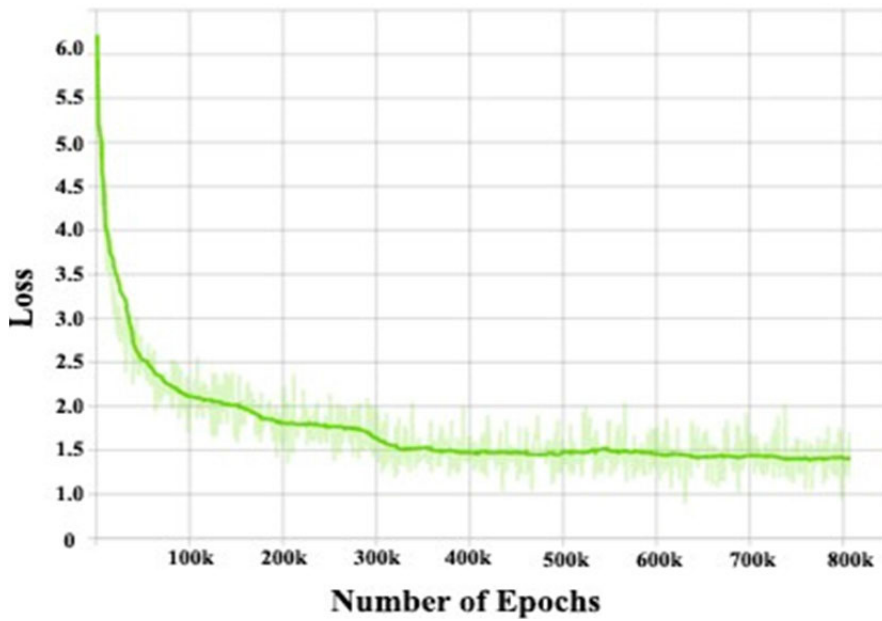


FIGURE 7. Mobile-Net average loss graph.



FIGURE 8. Practical analysis of prototype for a live video stream.

object detection (that is predicted correctly or incorrectly), the information related to the close match is conveyed to the user.

On the other hand, if the object is unknown, it will not be detected, and the user will get feedback that there is an unknown object and the system is unable to identify it. The object detection and recognition feature of the proposed system facilitates the user to search objects in the street, home, or store. It may also help the VIPs visualize the environment in which they are roaming. The second main feature of the prototype is to provide secure mobility to the user through an application created to monitor VIPs. Through this application, the family member can see the location of the user and

TABLE 2. Comparison of mobile net with other architectures.

Model	Image Net Accuracy	Million Parameters	Complexity (MPLOPs)	Classification error (%)
Mobile Net	71.8%	4.24	569	29.4
VGG 16	71.5%	138	15300	28.5
Squeezenet	57.5%	1.25	833	42.5
AlexNet	57.2%	60	720	42.8
GoogleNet	69.8%	6.8	1500	31.3
Shufflenet 2*(g=3)	73.7%	2	524	--

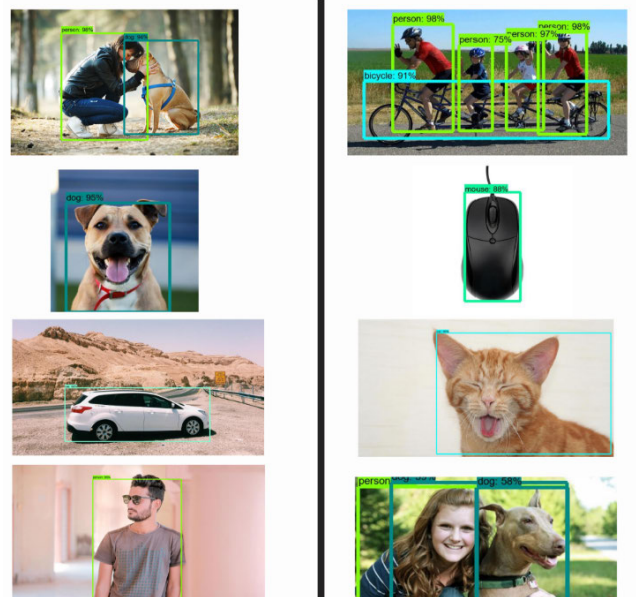


FIGURE 9. Practical analysis of prototype for offline images.

the snapshot/live feed. It is an on-demand/optional feature to ensure the user’s privacy. It is up to the user whether to turn it on or off. They may turn on this option only if they are lost.



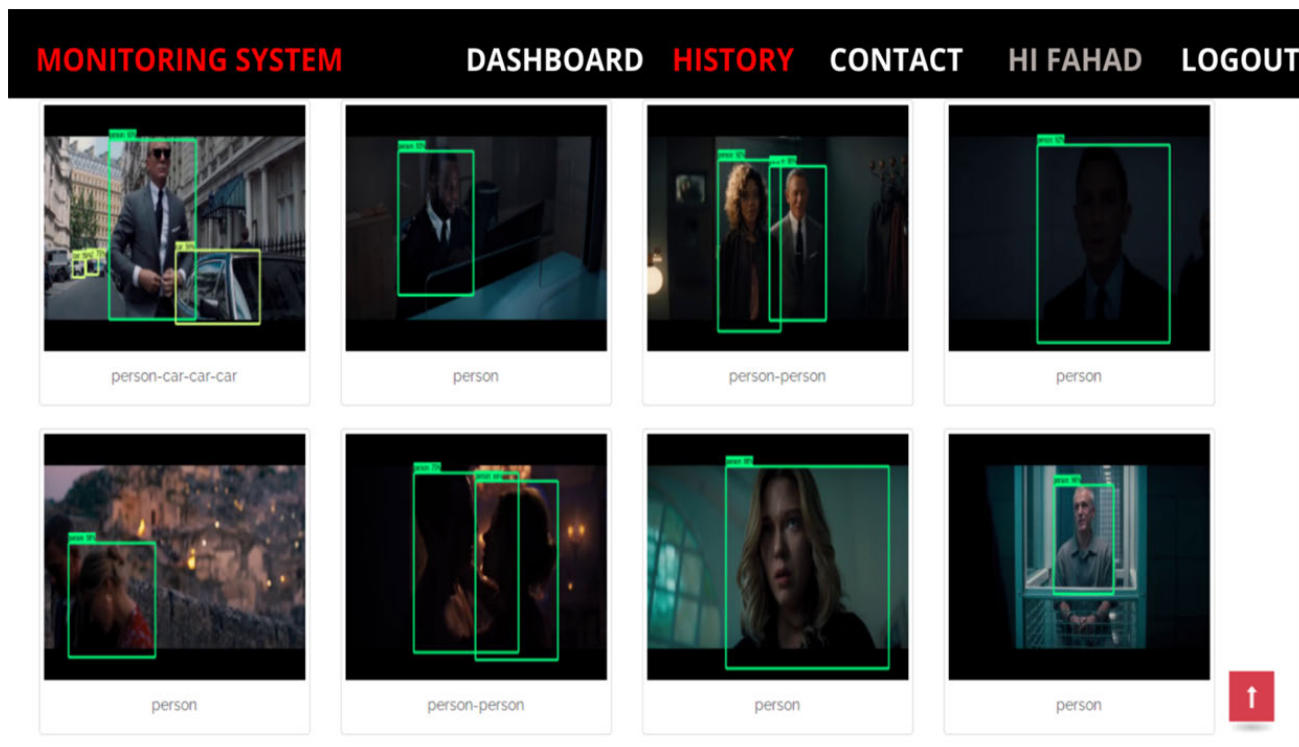


FIGURE 10. Practical analysis of prototype for offline videos.



FIGURE 11. Practical analysis of prototype for household objects.

After signing in the application, the observer is directed to the dashboard, which contains the most recent snapshots from the user and shows the user’s real-time location on Google map.

The prototype is tested on the most common object of day-to-day life in different environments to evaluate the

performance of the proposed system. In the first scenario, multiple objects from the live video stream are detected and recognized with great confidence, like a human, cat, horse, laptop, cup, bottle, etc. It also demonstrates the live detection of objects and shows the capability of deep learning

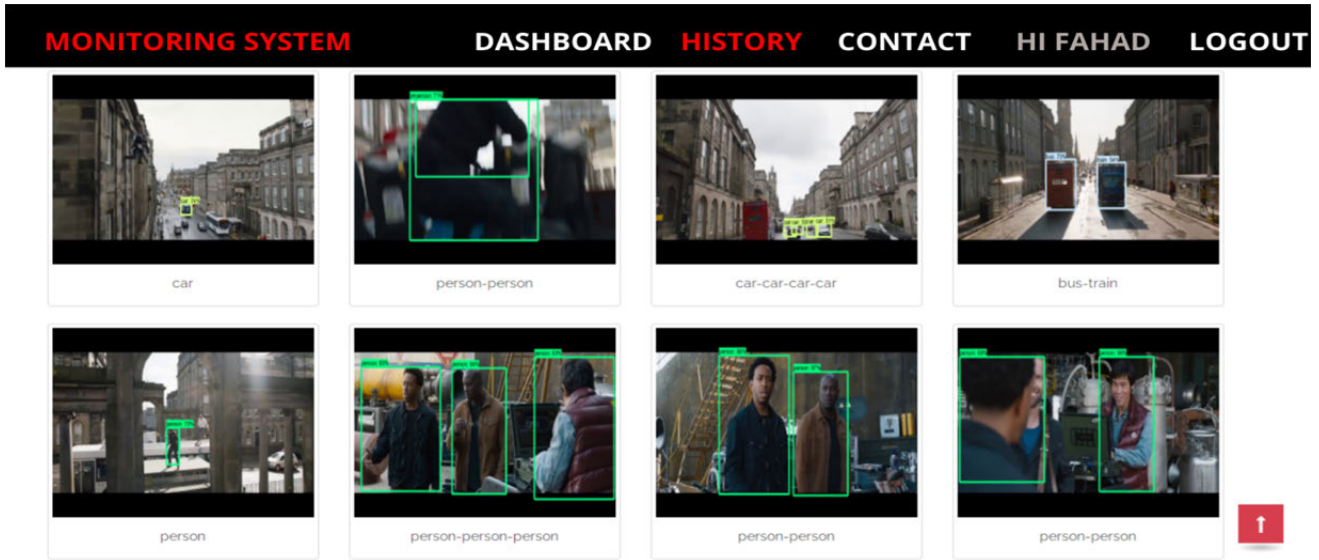


FIGURE 12. Practical analysis of prototype for the outdoor environment.

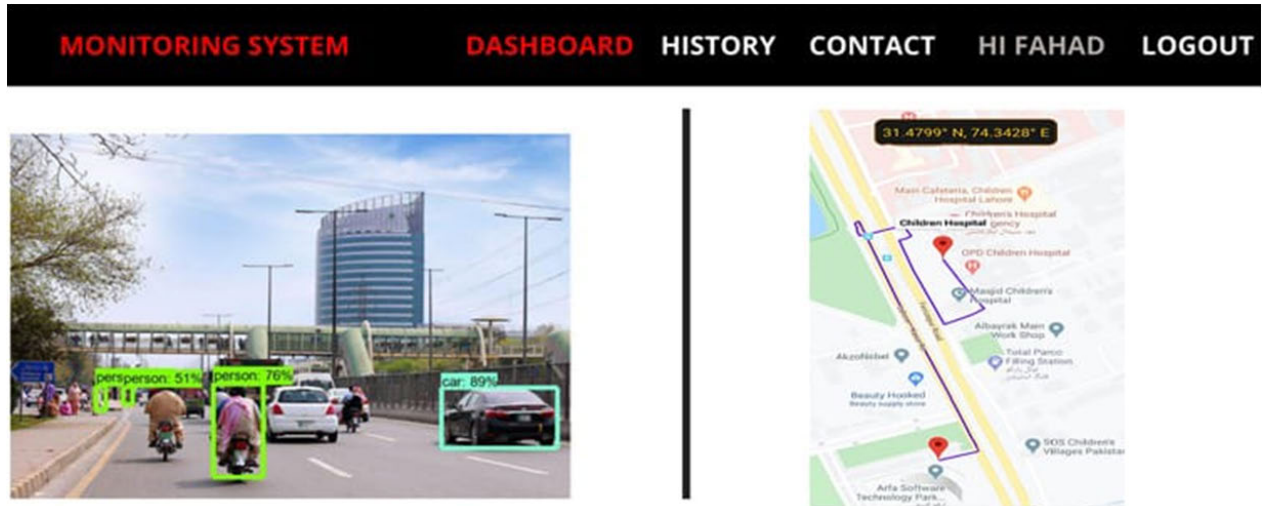


FIGURE 13. Live feedback and location showing on web-interface.

algorithms to localize and detect different objects. It also shows that the prototype can detect multiple objects in a single snapshot.

In the second scenario, the prototype is tested for offline images. The results are shown in Figure 9. Different images have been uploaded, and the system can easily recognize them accurately. The system can detect single or multiple objects. In frame 2, the prototype can detect multiple objects, such as persons with a maximum confidence level of 98% and bicycles with a 91% confidence level, even in the offline state.

In the third scenario, the prototype is tested for offline videos. The selection of videos is based on the principle that they should carry the potential disturbing phenomenon for object detection, e.g., bright scenes, dark scenes, etc. Also, the scenes should contain both indoor and outdoor environments. This is a pre-requisite condition to validate the flexibility and strength of the system to function properly in any situation.

Figure 10 shows the results for the different scenes of the test video. It is noticed that the system also works in dim lights. In the 1st frame, the person is moving outside. The prototype detects it with great confidence. In the 2nd frame, the system detects the person even in the dark. In the 3rd frame, the system easily recognizes the person with a 94% confidence level. It is also observed that the prototype functions properly for both (indoor, outdoor) environments.

In the fourth scenario, the prototype is tested for household objects. These are the most common objects of day-to-day life, and the system can easily detect and recognize them, as is displayed in Figure 11. In 1st frame, the system correctly detects the table and conveys the information to the user. Similarly, in the 4th frame, the system detects multiple objects in a single frame, i.e., the bed, the person sitting on the bed, and a vase.

In the fifth scenario, the prototype is tested in an outdoor environment. Figure 12 shows the experimental results for the





FIGURE 14. Real-time technical evaluation of device.

outdoor environment. In the 1st frame, the system detects the car, but it cannot detect multiple small-sized objects. In frame 4, there are two buses, and the system recognized one but was unable to recognize the other. Similarly, in other frames, the system provides accurate results.

In the sixth scenario, the prototype is tested in the outdoor environment to evaluate the second major feature, i.e., secure mobility. When this option is turned on, it shows the user's current location to the observer. The complete trajectory of the user is drawn on a Google map and can be monitored by the observer. In addition, the application periodically receives snapshots from the user to get details of the current environment. The working of this feature is shown in Figure 13.

**V. TECHNICAL EVALUATION OF REAL-TIME SCENARIOS**

The proposed system has been tested in real-time with four users (1 blind, 1 with covered eyes, two normal persons) in the age range of 20-25, as shown in figure 14.

The users used the prototype as displayed in figure 15. These users are instructed to move on multiple trajectories one by one and collect feedback regarding object detection and voice quality. In these trajectories, several objects were placed, e.g., dust bin, bottle, table, etc.

The results show that the system has correctly detected most objects along the path. The users could also listen to the object name in the current frame using the text to speech converter. The camera used to evaluate a system was able to detect obstacles within the range of 6m.

The result establishes that the system accurately detected obstacles and measured the distance of the hurdles. The system produces a fast response and allows a person to The objects comprise static (e.g., bed, table, and chair) and moving objects (e.g., bicycle, car, and pedestrian).

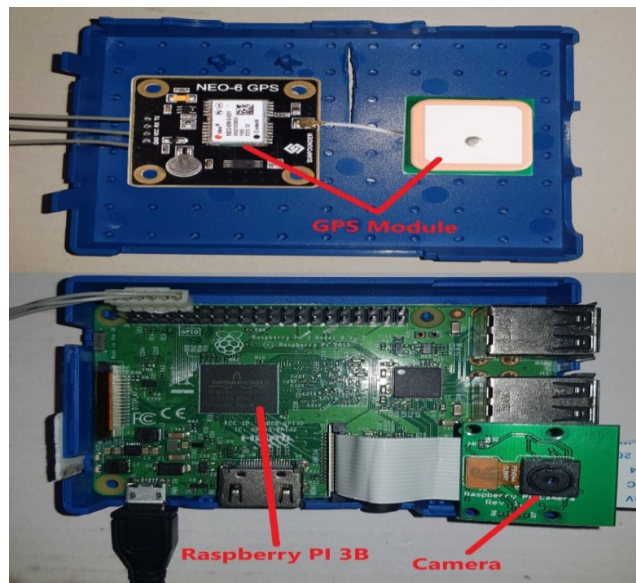


FIGURE 15. Pototype of the proposed device.

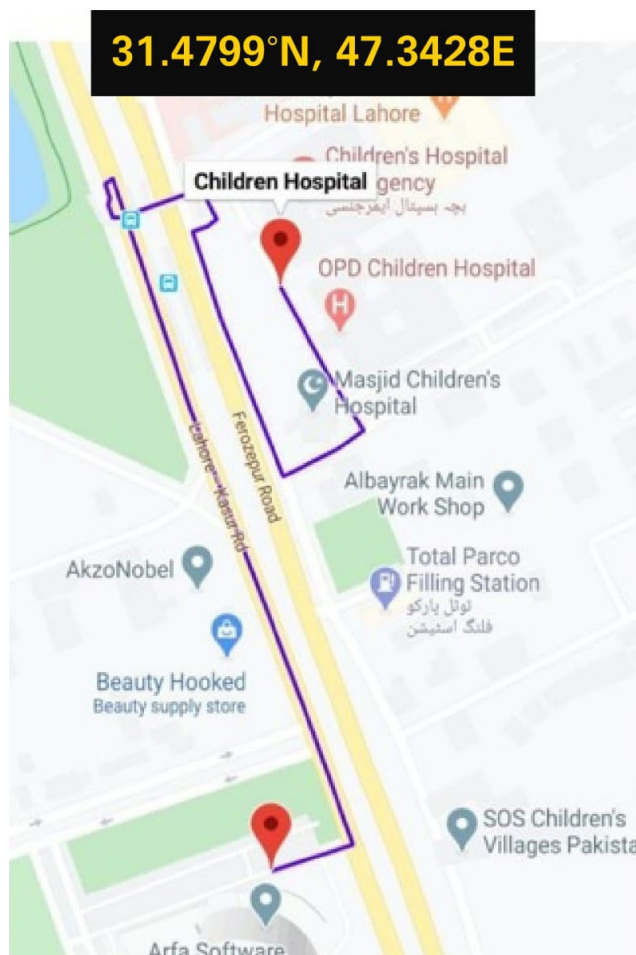


FIGURE 16. System current location and tracking.

A visually impaired person faces many problems in unfamiliar surroundings. An application has been created through Android studio integrated with Google Map API. The precise location and snapshot of the user were transmitted from the



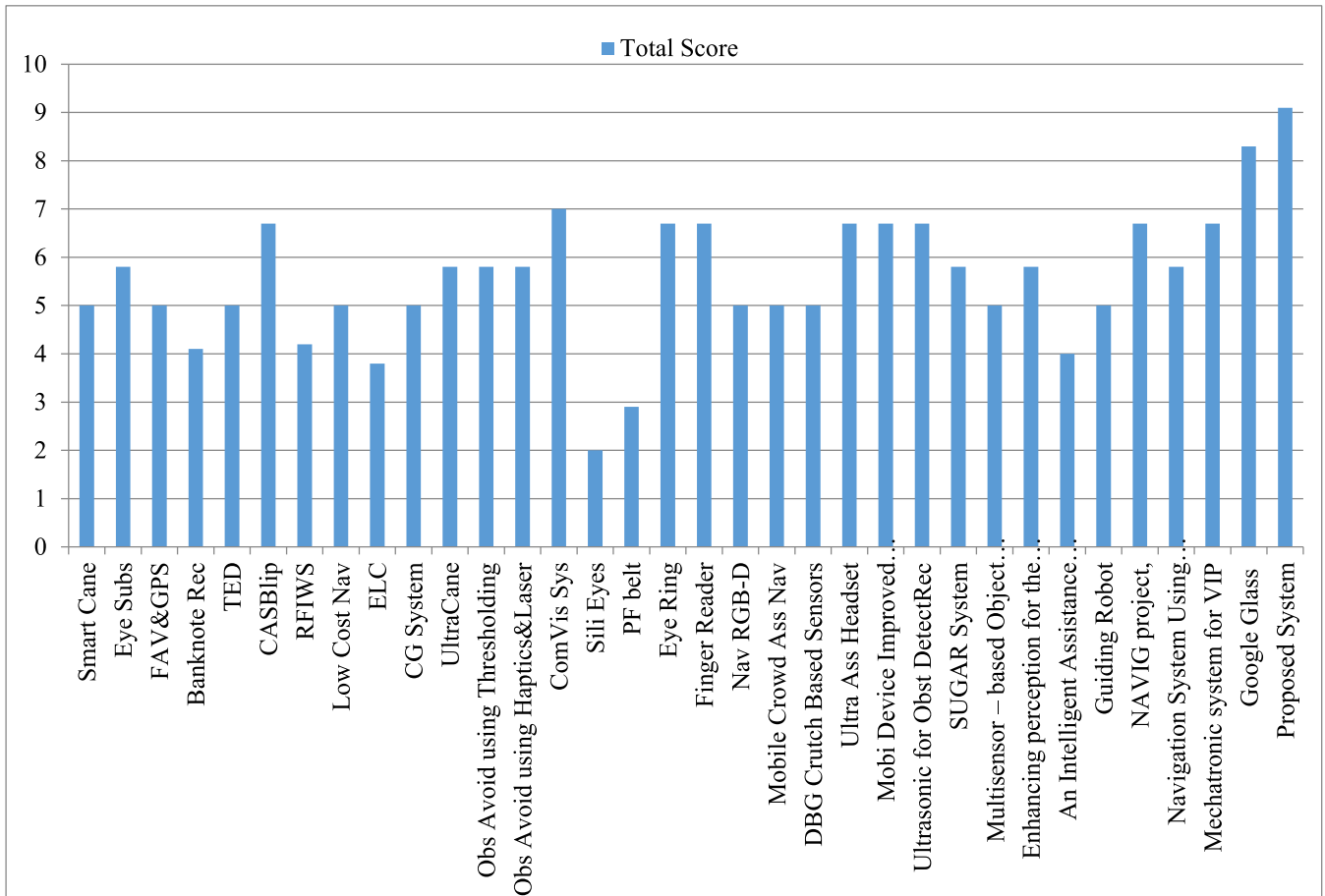


FIGURE 17. The quantitative analysis presents the total score of the systems.

application to the web server so that caregivers and family members could track them to ensure the security of VIPs. Several experiments are performed with moving user indoor and outdoor environments to evaluate the module. The application displayed the longitude and latitude of the location on the map, as shown in figure 16. The pilot testing of the system was successful as most of the members were able to navigate correctly.

The system also has a few limitations. During the experiment in the outdoor environment, it is noted that participants face problems in detecting road curbs, changes in the surface of the road, and detecting staircases. For in-door environment, participants faced difficulties in detecting soft fabric such as soft curtains, cushions, etc.

VI. COMPARATIVE ANALYSIS

This section presents the quantitative and comparative analysis for the proposed system with existing assistive devices. The quantitative analysis is based on the listed features in Table 3. There are a few essential attributes used by researchers in previous studies [1]. The assistive systems for VIPs need to contain the features such as concise and clear info within time, reliable performance throughout the day and night, proper working both in indoors and outdoors

TABLE 3. Features used in quantitative analysis.

Features	Description
Coverage Area	The device works in indoor/outdoor or both environments.
Time	The device functions at daytime/night or both.
Analysis Type	The response of the device in real-time or not.
Object Type	The device detects the static or dynamic objects
Security	The device helps in securing the person through monitoring.
Range	The device is either short range or long range.

environments, analysis in real-time, and a high accuracy rate. Otherwise, the manufacturer may not compete and survive in the market. The assessed features are essential for designing an assistive device/system for VIPs. For evaluation, each feature is assigned a weightage, i.e., if the developed assistive device contains all the mentioned features, its weightage is 10. Let us suppose that some devices work only in the daytime and vice versa, or some devices only work in indoor atmospheres and vice versa, then their score will be 5.

The assistive devices developed so far for the VIPs provide multiple features such as efficient performance during night or day times, working both in the outdoor and indoor environment, detecting the object, and identifying the dynamic

**TABLE 4. Score based quantitative comparative analysis.**

Devices	Features						Total Score
	Real time /Not Real time	Cover indoor / Outdoor, both	Time (day/night)	Range ( $R \leq 1$ m, $1 m < R \leq 5$ m, $R > 5$ m)	Security Rate	Object Type (Static, Dynamic, both)	
Weight of 10.0							
Smart Cane [49]	10	5	5	5	--	5	5
Eye Subs [17]	10	5	10	5	--	5	5.8
FAV&GPS [50]	10	5	5	--	--	10	5
Banknote Rec [51]	10	5	5	--	--	5	4.1
TED [15]	10	5	10		--	5	5
CASBlip [52]	10	10	10	5	--	5	6.7
RFIWS [53]	--	5	10	5	--	5	4.2
Low Cost Nav [16]	10	5	10	--	--	5	5
ELC [54]	10	5	10	2.5	--	5	3.8
CG System [55]	10	5	5	5	--	5	5
UltraCane [8]	10	5	10	5	--	5	5.8
Obs Avoid using Thresholding [6]	10	5	5	5	--	10	5.8
Obs Avoid using Haptics&Laser [56]	10	5	5	10	--	5	5.8
ComVis Sys [22]	10	10	5	10	--	10	7
Sili Eyes [57]	--	5	--	5	--	5	2
PF belt [58]	--	5	--	2.5	--	10	2.9
Eye Ring [59]	10	10	5	Specific Case 10	--	5	6.7
Finger Reader [59]	10	10	5	Specific Case 10	--	5	6.7
Nav RGB-D [19]	10	5	5	5	--	5	5
Mobile Crowd Ass Nav [60]	10	5	10		--	5	5
DBG Crutch Based Sensors [7]	10	5	5	5	--	5	5
Ultra Ass Headset [61]	10	10	10	5	--	5	6.7
Mobi Device Improved Verticle Resolution [62]	10	5	5	10	--	10	6.7
Ultrasonic for ObstDetectRec [63]	10	10	5	5	--	10	6.7
SUGAR System [64]	10	5	5	10	--	5	5.8
Multisensor – based Object Detection System [9]	10	5	--	10	--	5	5
Enhancing perception for the visually impaired with deep learning techniques. [10]	10	10	5	10	--	--	5.8
An Intelligent Assistance System for Visually Impaired [11]	10	5	--	10	--	--	4
Guiding Robot [31]	10	--	--	10	--	10	5
NAVIG project, [36]	10	10	10	--	--	10	6.7
Navigation System Using HoloLens and RFID [20]	10	10	5	--	--	10	5.8
Mechatronic system for VIP [21]	10	10	10	--	--	10	6.7
Google Glass [65]	10	10	10	10	--	10	8.3
<b>Proposed System</b>	<b>10</b>	<b>10</b>	<b>5</b>	<b>10</b>	<b>10</b>	<b>10</b>	<b>9.1</b>

or static objects. A few of the devices assist in navigation and cover many other features. The evaluated features are the basic features that assist in designing assistive devices.

The highest weightage is 10, achieved by devices having all the basic features. The score for each feature in every system is highlighted depending on the collected information.

Some of the features are still in the research stage. The user's response is also considered to evaluate only those devices tested in real-time scenarios. The value of a device is considered as  $V_k$ . The range is between 0-10, where the '10' value is given to the device that contains all features. The feature value of 5 is given to the system that will perform in only indoor conditions, or the device works only in the outdoor environment, e.g., Smart Cane and other devices. The same criteria are being applied to the range of the device feature. The '10' value is assigned if it covers a large range. The value 2.5 would be assigned to those devices whose detection range is low i.e., less than or equal to 1m. An equation is defined to compute the total score of every system based on the features in Table 2. It is as follows,

$$\text{Total Score} = \sum_{k=1}^N \frac{V_k}{N} \quad (2)$$

where  $N$  represents the total number of features and  $k$  describes a particular feature.

Table 4. gives a quantitative analysis of the assistive devices. It is evident from Table 4 that the proposed system outperformed others in terms of the total score. It is due to the enriched features supported by the proposed system. It is important to note that none of the devices is perfect, and they need improvements in their design to fulfill the requirements of the VIPs.

Figure 17 presents a complete picture of the evaluation for each system with an individual score. Systems with higher scores demonstrate robust and enhanced performance.

## VII. CONCLUSION

This paper presented a smart and intelligent system for VIPs to assist them in mobility and ensure their safety. The proposed system is based on the day-to-day requirements of VIPs. It assists them in visualizing the environment and providing a sense of the surroundings. They can recognize objects around them and sense the natural environment using CNN-based low-power Mobile-Net architecture. Moreover, a web-based application is developed to ensure the safety of VIPs. The user of this application can turn the on-demand function to share his/her location with the family. It is beneficial to their family as they can monitor the movement of VIPs and can track his/her location using the live feed from the camera. The experimental analysis shows that the proposed system provided satisfactory results and outperformed other devices in terms of supported features.

The task is very challenging but using state-of-the-art techniques, the suitability of a device can be computed. In the future, additional criteria for selecting features, such as aesthetic sense and outlook of the device, may also be considered to evaluate its performance. Moreover, the bone conduction headphone may also be incorporated to improve feedback. Finally, after launching the device, surveys and user comments will also be gathered to enhance the performance of a device.

## REFERENCES

- [1] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, p. 565, 2017.
- [2] *Work Sheet*. Accessed: Mar. 7, 2020. [Online]. Available: <https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [3] R. Velázquez, "Wearable assistive devices for the blind," in *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment*. Berlin, Germany: Springer, 2010.
- [4] L. B. Neto, F. Grijalva, V. R. M. L. Maike, L. C. Martini, D. Florencio, M. C. C. Baranauskas, A. Rocha, and S. Goldenstein, "A Kinect-based wearable face recognition system to aid visually impaired users," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 52–64, Feb. 2017.
- [5] C. Shah, M. Bouzid, M. Youssef, and L. Vasquez, "Evaluation of RU-netractable feedback navigation system for the visually impaired," in *Proc. Int. Workshop Virtual Rehabil.*, 2006, pp. 72–77.
- [6] M. R. U. Saputra, Widyawan, and P. I. Santosa, "Obstacle avoidance for visually impaired using auto-adaptive thresholding on Kinect's depth image," in *Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput., IEEE 11th Int. Conf. Auton. Trusted Comput., IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops*, Dec. 2014, pp. 337–342.
- [7] Y. Yi and L. Dong, "A design of blind-guide crutch based on multi-sensors," in *Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, Aug. 2015, pp. 2288–2292.
- [8] K. Kumar, B. Champaty, K. Uvanesh, R. Chachan, K. Pal, and A. Anis, "Development of an ultrasonic cane as a navigation aid for the blind people," in *Proc. Int. Conf. Control, Instrum., Commun. Comput. Technol. (ICCICCT)*, Jul. 2014, pp. 475–479.
- [9] C. T. Patel, V. J. Mistry, L. S. Desai, and Y. K. Meghrajani, "Multisensor-based object detection in indoor environment for visually impaired people," in *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2018, pp. 1–4.
- [10] Z. Bauer, A. Dominguez, E. Cruz, F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors," *Pattern Recognit. Lett.*, vol. 137, pp. 27–36, Sep. 2020.
- [11] L.-B. Chen, J.-P. Su, M.-C. Chen, W.-J. Chang, C.-H. Yang, and C.-Y. Sie, "An implementation of an intelligent assistance system for visually impaired/blind people," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp. 1–2.
- [12] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2016, pp. 208–213.
- [13] S.-H. Chae, M.-C. Kang, J.-Y. Sun, B.-S. Kim, and S.-J. Ko, "Collision detection method using image segmentation for the visually impaired," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 392–400, Nov. 2017.
- [14] P. Salavati and H. M. Mohammadi, "Obstacle detection using GoogleNet," in *Proc. 8th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2018, pp. 326–332.
- [15] T. H. Nguyen, T. L. Le, T. T. H. Tran, N. Vuillermé, and T. P. Vuong, "Antenna design for tongue electro-tactile assistive device for the blind and visually-impaired," in *Proc. 7th Eur. Conf. Antennas Propag. (EuCAP)*, 2013, pp. 1183–1186.
- [16] J. Xiao, K. Ramdath, M. Iosilevish, D. Sigh, and A. Tsakas, "A low cost outdoor assistive navigation system for blind people," in *Proc. IEEE 8th Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2013, pp. 828–833.
- [17] S. Bharambe, R. Thakker, H. Patil, and K. M. Bhurchandi, "Substitute eyes for blind with navigator using android," in *Proc. Texas Instrum. India Educ. Conf.*, Apr. 2013, pp. 38–43.
- [18] A. S. Martínez-Sala, F. Losilla, J. C. Sánchez-Aarnoutse, and J. García-Haro, "Design, implementation and evaluation of an indoor navigation system for visually impaired people," *Sensors*, vol. 15, pp. 32168–32187, Dec. 2015.
- [19] A. Aladrén, G. López-Nicolás, L. Puig, and J. J. Guerrero, "Navigation assistance for the visually impaired using RGB-D sensor with range expansion," *IEEE Syst. J.*, vol. 10, no. 3, pp. 922–932, Sep. 2016.
- [20] A. Yamashita, K. Sato, S. Sato, and K. Matsubayashi, "Pedestrian navigation system for visually impaired people using HoloLens and RFID," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Dec. 2017, pp. 130–135.
- [21] A. Mancini, E. Frontoni, and P. Zingaretti, "Mechatronic system to help visually impaired users during walking and running," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 649–660, Feb. 2018.



- [22] S. Caraiman, A. Morar, M. Owczarek, A. Burlacu, D. Rzeszotarski, N. Botezatu, P. Herghelegiu, F. Moldoveanu, P. Strumillo, and A. Moldoveanu, "Computer vision for the visually impaired: The sound of vision system," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 1480–1489.
- [23] R. Jiang, Q. Lin, and S. Qu, "Let blind people see: Real-time visual recognition with results converted to 3D audio," Standord Univ., Stanford, CA, USA, Tech. Rep. 218, 2016.
- [24] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A real-time object detection algorithm optimized for non-GPU computers," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2503–2510.
- [25] K. Yang, K. Wang, W. Hu, and J. Bai, "Expanding the detection of traversable area with real sense for the visually impaired," *Sensors*, vol. 16, no. 11, p. 1954, 2016.
- [26] Q. Zhao, B. Zhang, S. Lyu, H. Zhang, D. Sun, G. Li, and W. Feng, "A CNN-SIFT hybrid pedestrian navigation method based on first-person vision," *Remote Sens.*, vol. 10, no. 8, p. 1229, 2018.
- [27] A. Kunz, K. Miesenberger, L. Zeng, and G. Weber, "Virtual navigation environment for blind and low vision people," in *Proc. Int. Conf. Comput. Helping People With Special Needs*, 2018, pp. 114–122.
- [28] X. Zhang, X. Yao, Y. Zhu, and F. Hu, "An ARCore based user centric assistive navigation system for visually impaired people," *Appl. Sci.*, vol. 9, no. 5, p. 989, Mar. 2019.
- [29] M. Kubanek and J. Bobulski, "Device for acoustic support of orientation in the surroundings for blind people," *Sensors*, vol. 18, no. 12, p. 4309, Dec. 2018.
- [30] M. M. Rahman, M. M. Islam, and S. Ahmmed, "'BlindShoe': An electronic guidance system for the visually impaired people," *J. Telecommun. Electron. Comput. Eng.*, vol. 11, no. 2, pp. 49–54, 2019.
- [31] R. K. Megalingam, S. Vishnu, V. Sasikumar, and S. Sreekumar, "Autonomous path guiding robot for visually impaired people," in *Cognitive Informatics and Soft Computing*. Singapore: Springer, 2019, pp. 257–266.
- [32] M. M. Rahman, M. M. Islam, S. Ahmmed, and S. A. Khan, "Obstacle and fall detection to guide the visually impaired people with real time monitoring," *Social Netw. Comput. Sci.*, vol. 1, pp. 1–10, Jul. 2020.
- [33] R. A. Minhas and A. Javed, "X-EYE: A bio-smart secure navigation framework for visually impaired people," in *Proc. Int. Conf. Signal Process. Inf. Secur. (ICSPIS)*, Nov. 2018, pp. 1–4.
- [34] V. V. Meshram, K. Patil, V. A. Meshram, and F. C. Shu, "An astute assistive device for mobility and object recognition for visually impaired people," *IEEE Trans. Human-Machine Syst.*, vol. 49, no. 5, pp. 449–460, Oct. 2019.
- [35] A. Berger, A. Vokalova, F. Maly, and P. Poulouva, "Google glass used as assistive technology its utilization for blind and visually impaired people," in *Proc. Int. Conf. Mobile Web Inf. Syst.*, 2017, pp. 70–82.
- [36] S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, A. Serpa, M. Raynal, B. Oriola, M. M. Macé, M. Auvray, M. Denis, and S. J. Thorpe, "Navigation and space perception assistance for the visually impaired: The NAVIG project," *IRBM*, vol. 33, no. 2, pp. 182–189, Apr. 2012.
- [37] W.-J. Chang, L.-B. Chen, C.-Y. Sie, and C.-H. Yang, "An artificial intelligence edge computing-based assistive system for visually impaired pedestrian safety at zebra crossings," *IEEE Trans. Consum. Electron.*, vol. 67, no. 1, pp. 3–11, Feb. 2021.
- [38] W.-J. Chang, L.-B. Chen, M.-C. Chen, J.-P. Su, C.-Y. Sie, and C.-H. Yang, "Design and implementation of an intelligent assistive system for visually impaired people for aerial obstacle avoidance and fall detection," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10199–10210, Sep. 2020.
- [39] M. A. Rahman and M. S. Sadi, "IoT enabled automated object recognition for the visually impaired," *Comput. Methods Programs Biomed. Update*, vol. 1, May 2021, Art. no. 100015.
- [40] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognit. Lett.*, vol. 137, pp. 37–52, Sep. 2020.
- [41] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [42] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
- [43] *Speech API Overview (SAPI 5.3)*. Accessed: Apr. 3, 2020. [Online]. Available: [https://docs.microsoft.com/en-us/previousversions/windows/desktop/ms720151\(v%3Dvs.85\)](https://docs.microsoft.com/en-us/previousversions/windows/desktop/ms720151(v%3Dvs.85))
- [44] *Raspberry Pi 3G/4G & LTE Base HAT*. Accessed: Jul. 26, 2021. [Online]. Available: <https://sixfab.com/product/raspberry-pi-base-hat-3g-4g-lte-minipiccards/>
- [45] *Django*. Accessed: Jan. 21, 2020. [Online]. Available: <https://www.djangoproject.com/>
- [46] *Image Net*. Accessed: Apr. 3, 2020. [Online]. Available: <https://cloud.google.com/tpu/docs/imagenet-setup>
- [47] *Mobile Net*. Accessed: Apr. 4, 2020. [Online]. Available: <https://github.com/Zehaos/MobileNet>
- [48] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
- [49] M. H. A. Wahab, A. A. Talib, H. A. Kadir, A. Johari, A. Noraziah, R. M. Sidek, and A. A. Mutalib, "Smart cane: Assistive cane for visually-impaired people," 2011, *arXiv:1110.5156*.
- [50] A. Brilhault, S. Kammoun, O. Gutierrez, P. Truillet, and C. Jouffrais, "Fusion of artificial vision and GPS to improve blind pedestrian positioning," in *Proc. 4th IFIP Int. Conf. New Technol., Mobility Secur.*, Feb. 2011, pp. 1–5.
- [51] A. Sirikhram, W. Chirachrit, and K. Chamnongthai, "Banknote and coin speaker device for blind people," in *Proc. 11th Int. Conf. Adv. Commun. Technol.*, vol. 3, 2009, pp. 2137–2140.
- [52] L. Dunai, B. D. Garcia, I. Lengua, and G. Peris-Fajarnes, "3D CMOS sensor based acoustic object detection and navigation system for blind people," in *Proc. 38th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2012, pp. 4208–4215.
- [53] M. F. Saaid, I. Ismail, and M. Z. H. Noor, "Radio frequency identification walking stick (RFIWS): A device for the blind," in *Proc. 5th Int. Colloq. Signal Process. Appl.*, Mar. 2009, pp. 250–253.
- [54] A. R. García, R. Fonseca, and A. Durán, "Electronic long cane for locomotion improving on visual impaired people. A case study," in *Proc. Pan Amer. Health Care Exchanges*, 2011, pp. 58–61.
- [55] A. Landa-Hernández and E. Bayro-Corrochano, "Cognitive guidance system for the blind," in *Proc. World Autom. Congr.*, 2012, pp. 1–6.
- [56] D. I. Ahlmark, H. Fredriksson, and K. Hyppä, "Obstacle avoidance using haptics and a laser rangefinder," in *Proc. IEEE Workshop Adv. Robot. Social Impacts*, Nov. 2013, pp. 76–81.
- [57] B. R. Prudhvi and R. Bagani, "Silicon eyes: GPS-GSM based navigation assistant for visually impaired using capacitive touch Braille keypad and smart SMS facility," in *Proc. World Congr. Comput. Inf. Technol. (WCCIT)*, Jun. 2013, pp. 1–3.
- [58] J. F. Oliveira, "The path force feedback belt," in *Proc. 8th Int. Conf. Inf. Technol. Asia (CITA)*, Jul. 2013, pp. 1–6.
- [59] R. Shilkrot, J. Huber, C. Liu, P. Maes, and S. C. Nanayakkara, "Finger-Reader: A wearable device to support text reading on the go," in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2014, pp. 2359–2364.
- [60] G. Olmschenk, C. Yang, Z. Zhu, H. Tong, and W. H. Seiple, "Mobile crowd assisted navigation for the visually impaired," in *Proc. IEEE 12th Int. Conf. Ubiquitous Intell. Comput., IEEE 15th Int. Conf. Scalable Comput. Commun. Associated Workshops (UIC-ATC-ScalCom)*, Aug. 2015, pp. 324–327.
- [61] S. Aymaz and T. Cavdar, "Ultrasonic assistive headset for visually impaired people," in *Proc. 39th Int. Conf. Telecommun. Signal Process. (TSP)*, Jun. 2016, pp. 388–391.
- [62] L. Everding, L. Walger, V. S. Ghaderi, and J. Conradt, "A mobility device for the blind with improved vertical resolution using dynamic vision sensors," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2016, pp. 1–5.
- [63] B. Mocanu, R. Tapu, and T. Zaharia, "When ultrasonic sensors and computer vision join forces for efficient obstacle detection and recognition," *Sensors*, vol. 16, no. 11, p. 1807, 2016.
- [64] S. Duman, A. Elewi, and Z. Yetgin, "Design and implementation of an embedded real-time system for guiding visually impaired individuals," in *Proc. Int. Artif. Intell. Data Process. Symp. (IDAP)*, Sep. 2019, pp. 1–5.
- [65] A. Berger, A. Vokalova, F. Maly, P. Poulouva, W. Elmannai, and K. Elleithy, "Google glass used as assistive technology its utilization for blind and visually impaired people," *Sensors*, vol. 17, no. 3, p. 565, 2017.



**FAHAD ASHIQ** received the B.S. degree in computer science from Lahore Leads University, Lahore, Pakistan, in 2018. He is currently pursuing the master's degree in information engineering and computer science from Rhine-Waal University of Applied Sciences, Germany. His research interests include image and video processing, computer vision, machine learning, and web-engineering.



optimization, and network security.

**MUHAMMAD ASIF** received the Ph.D. degree in electrical engineering from the Capital University of Science and Technology (CUST), Islamabad, Pakistan, in 2016. He is currently an Assistant Professor with the Department of Computer Science, Lahore Garrison University (LGU), Lahore, Pakistan. He has contributed more than 40 research papers. His research interests include image and video processing, computer vision, machine learning, parallel processing, embedded systems, optimization, and network security.



**MAAZ BIN AHMAD** received the Ph.D. degree in computer engineering from the Centre for Advanced Studies in Engineering (CASE), Islamabad, Pakistan, in 2014. He is currently an Assistant Professor with the College of Computing and Information Sciences, Karachi Institute of Economics and Technology, Karachi, Pakistan. He has published more than 25 research papers. His research interests include network security, video processing, and multimedia.



**SADIA ZAFAR** received the M.S. degree in computer sciences from Lahore Garrison University (LGU), Lahore, Pakistan, in 2020. She is currently working as a Lecturer in Software Engineering Department, LGU. She has contributed in several research articles. Her research interests include image and video processing, computer vision and machine learning, and web engineering.



ing research. He has authored a number of reputed journal publications and top tier computer science conferences. His research interests include computer vision, machine learning, medical imaging, and biomedical signal processing.

**KHALID MASOOD** received the Ph.D. degree in computer science from the University of Warwick, U.K., in 2010. He is currently working as an Associate Professor with Lahore Garrison University, Pakistan. He is actively involved in the processing of biomedical signals and applications of deep learning in health care and cancer research. He holds a patent (filed in the USA) for the cancer research and has been awarded two grants for the medical imaging and biomedical signal processing research. He has authored a number of reputed journal publications and top tier computer science conferences. His research interests include computer vision, machine learning, medical imaging, and biomedical signal processing.



inferences and journals of international repute. His research interests include image processing, image retrieval, steganography, data science, and numerical techniques with particular attention to multimedia forensics. He is serving as an Academic Editor for *Mathematical Problems in Engineering* (Hindawi). He has been serving as a Reviewer for *The Visual Computer*, *ETRI Journal*, *Journal of Information Security and Applications*, *Australian Journal of Forensic Sciences*, *Forensic Science International*, *IEEE Access*, *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO*, *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *Journal of Internet Technology*, and many more.

**TOQEER MAHMOOD** received the M.S. degree in computer engineering from the Center for Advanced Studies in Engineering (CASE), Islamabad, Pakistan, in 2010, and the Ph.D. degree in computer engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2017. He is currently an Assistant Professor of computer science and engineering at the National Textile University, Faisalabad, Pakistan. He has authored or coauthored many scientific papers



of Computer Science and Engineering, Korea University of Technology and Education (Koreatech). His research interests include 3-D shape recovery from image focus, computer vision, pattern recognition, and machine learning.

**MUHAMMAD TARIQ MAHMOOD** (Senior Member, IEEE) received the M.C.S. degree in computer science from the AJK University of Muzaffarabad, Pakistan, in 2004, the M.S. degree in intelligent software systems from the Blekinge Institute of Technology, Sweden, in 2006, and the Ph.D. degree in information and mechatronics from the Gwangju Institute of Science and Technology, South Korea, in 2011. He is currently working as an Associate Professor at the School



Assistant Professor with the Department of Mechatronics Engineering and Korea Polytechnic University, South Korea. His research interests include image registration, image fusion, depth estimation, and medical image processing.

**IK HYUN LEE** received the B.S. degree in control and instrument engineering from Korea University, South Korea, in 2004, and the M.S. and Ph.D. degrees from the School of Information and Mechatronics, Gwangju Institute of Science and Technology, South Korea, in 2008 and 2013, respectively. He was a Postdoctoral Researcher at the Media Laboratory, Massachusetts Institute of Technology, and a Senior Researcher at the Korea Aerospace Institute of Research. He is currently an

...