Received December 29, 2021, accepted January 17, 2022, date of publication January 25, 2022, date of current version February 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3146320

Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification

USMAN MASUD[1,2], TAREQ SAEED[3], HUNIDA M. MALAIKAH[4], FEZAN UL ISLAM[1], AND GHULAM ABBAS[1]

[1]Department of Electrical Communication Engineering, University of Engineering and Technology, Taxila 47050, Pakistan
[2]Department of Electrical Communication Engineering, University of Kassel, 34127 Kassel, Germany
[3]Nonlinear Analysis and Applied Mathematics (NAAM)-Research Group, Department of Mathematics, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[4]Department of Mathematics, Faculty of Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Usman Masud (usmanmasud123@hotmail.com)

This work was supported in part by the German Academic Exchange Service [Deutsche Akademische Austausch Dienst (DAAD)], and in part by the EU.

ABSTRACT Recent progress in innovation is making the life prosper, simpler and easier for common individual. The World Health Organization (WHO) statistics indicate that a large amount of people experience visual losses, because of which they encounter many difficulties in everyday jobs. Hence, our goal is to structure a modest, secure, wearable, and versatile framework for visually impaired to help them in their daily routines. For this, the plan is to make an effective system which will assist visually impaired people through obstacle detection and scenes classification. The proposed methodology utilizes Raspberry-Pi 4B, Camera, Ultrasonic Sensor and Arduino, mounted on the stick of the individual. We take pictures of the scene and afterwards pre-process these pictures with the help of Viola Jones and TensorFlow Object Detection algorithm. The said techniques are used to detect objects. We also used an ultrasonic sensor mounted on a servomotor to measure the distance between the blind person and obstacles. The presented research utilizes simple calculations for its execution, and detects the obstructions with a notably high efficiency. When contrasted with different frameworks, this framework is a minimal effort, convenient, and simple to wear.

INDEX TERMS Smart system, visual losses, biomedical sensor, object recognition, tensorflow, Viola Jones, ultrasonic sensor.

I. INTRODUCTION

Life of any individual depends on basic five senses in which ability of vision is probably the most important one. Visual impairment is a decreased ability to see something to a level that the eye is not able to see even using usual means, such as lenses or glasses. Visually impaired individuals [1] lack this sense of vision. Hence, fulfilling the daily tasks of life becomes extremely hard for them. This can lead to difficulties which can only be temporarily subdued by some assisting personnel, and cases exist where certain situations might be fatal, not only for the individual, but also anyone in the surrounding environment [2]. With the innovations in scientific area, much research was proposed that specifies the designing of gadgets for visually impaired individuals. These gadgets [2]–[4] were simple and durable, but they were deficient in terms of usage and accuracy.

As the modern world relies on computer and artificial intelligence, these techniques became more reliable and efficient. However, many gaps remain in these technologies. During our study to previous literature, we found that RFID sensors, Logitech cameras, embedded systems [5] were used in past to design an efficient system. We know that as the visual impairments diseases are increasing with the passage of time. Therefore, our motivation of this project is to make a system which will assist visually defected people by classifying scenes and avoiding obstacles through object detection, to lead a life like other normal people in world. If there is an obstacle in the user's path, the system will track it, and convey that user about it.

During this process, there are numerous obstacles that can appear across the length of the distance [5], [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

TABLE 1. List of symbols.

WHO	World Health Organization
RFID	Radio Frequency Identification
ETA	Electronic Travel Aids
EOA	Electronic Orientation Aids
GSM	Global System for Mobile communication
GPS	Global Positioning System
SONAR	Sound Navigation and Ranging
IDE	Integrated Development Environment fluctuation
ANN	Artificial Neural Network
COCO	Common Objects in Context (Dataset)
VNC	Virtual Network Computing
SLAM	Simultaneous Localization and Mapping
NDLF	National Digital Library Federation
IOT	Internet of Things
PDR	Preliminary Data Requirements
PLW	Point Locus Wearable

This information is important for the user. For instance, there is a stool across the path of the user and must be recognized. Otherwise, the user can be harmed. This obstacle must be recognized in time, in order that it can be avoided afterwards [7]–[10]. On the contrary, if there is a something small as a pebble, this can be ignored by the system. The reason being that normal persons also ignore small things that do not intend to cause any significant harm. Therefore, the system will also recognize the scenes [2], [11], [12] for the intended user. All this information will be communicated in the form of voice to the person, in order to avoid any damage.

II. RELATED WORK

Several techniques are used for the navigation of visually impaired individuals by health departments, for instance, Electronic Orientation Aids (EOAs), Place Locator Devices (PLDs) and Electronic Travel Aids (ETAs) [13]–[15]. The suggested technology not only provides individuals with directions, but it also monitors the health of visually impaired persons. A clever ETA is proposed in this study [13], where an ultrasonic sensor is used to detect and echo sound waves that can detect an obstruction at 5 to 35 cm. The suggested approach detects things at the ground and waist levels up to a maximum of 2 m using two specified placements of ultrasonic sensors. There is another approach that is based on non-intrusive wearable devices [16]. Here, three dimensional and semantic data is used to represent the scenes in front of the user, and the information is then transmitted to user through text to speech system or haptic feedback.

A small and lightweight transmission system that aids the blind in sign interpretation is presented in [17] in which wireless transmission across the visible light frequency range has been proposed for communication. To communicate between a sign and a wearable smart glass, a light transmission mechanism works in this manner. In another system [18], Internet of Things (IOT) has been used which is heavily reliant on data science and analytics. Data collected may be analyzed and utilized to identify impediments and improves basic navigation with haptic and vocal feedback [19]. Using this approach, identifying everyday things becomes less complicated.

Infrared energy cannot be seen with the naked eye, but it may be observed using special cameras that convert light into visible colors. Using this technique [20], infrared technology is utilized to improve vision impairment, utilizing mobile phones and an application that runs on an Android-based smartphones. This leads to *picture captioning* which refers to the process of creating a written explanation for an image. It employs Machine Learning techniques such as Natural Language. For our dataset photos, National Digital Library Federation (NDLF) [21] provides relevant and accurate captions. This technique has an accuracy of 82.39%.

Because of signal losses, a Global Positioning System (GPS) does not operate effectively indoors [22], [23]. To deal with such situations, there is a technique called Preliminary Data Requirements (PDR) that estimates a pedestrian's position using inertial sensors [2], [25]. Inertial sensors include motion sensors (accelerometers) and rotation sensors (gyroscope sensors) to uninterruptedly measure by dead reckoning the position, the orientation, and the velocity (direction and speed of movement) of a moving object. Using this technology, people with vision impairments may walk around securely indoors. As it uses GPS for outdoor navigations and PDR technique for indoor navigation, so it cannot identify the objects and assist the vision impairer properly. Thus, it cannot be implemented locally.

A deep learning approach [24] is used to evaluate safe and reliable procedure through the data od RGB images and establish a semantic map that helps in getting good results in both indoor and outdoor scenes. In a similar manner, the smart glass system [26], which uses a convolutional neural network to transmit auditory signal through headphones, can read the name of the drug. Because of the complexity and limited processing resources, visual object recognition still remains an unresolved topic.

Embedded systems are utilized in walking sticks by vision-impaired people in another technique [27], where notification is done by speech over headphones or speakers, with a voice recorder and playback IC. Raspberry Pi 4 was used, which has improved computational capability and is connected to the pi camera, GPS, and GSM modules. With the aid of GPS, this gadget [28] tracks the user's present location. When impediments are identified in the path, it also emits a warning sound.

In another work [29], Zhe-Wang gathered inside course structures for astonished customers that use RGB (Red, Green, Blue) and depth cameras to get consistent pictures. The collected images were then pre-handled, *utilizing multi scaling technology*, and the yield was delivered to the buyer through speaker. This structure gives a precision of 73.6

RGB profundity cameras are used by Tian to recognize steps including walker crosslines [30], [31]. Hough Transformation calculation was used for setting up the pictures and afterward gathering is finished utilizing the SVM and achieved an accuracy of 93.90

In another investigation, James [32] created a scene enhancement structure that employs the Simultaneous Localization and Mapping method (SLAM) to detect deterrents. Camera and glasses transfer images to a PDA [[34], [36], which then sends these images to the client through earphone. Investigators create a PC vision-based assistance for seemingly disabled people that employs a Logitech camera for image collection [38]. The images were then processed using the Hough Transform computation, and the removed features were requested using the Fuzzy Logic Classifier.

A multi-object detection technique uses Artificial Intelligence and smart navigation for visually Impaired people in which multiple images of objects that are highly relevant to the visually impaired people are used to train a deep learning model [37]. Point Locus Wearable GPS Pathfinder System has been intended to assist individuals with visual disabilities to travel outside [39]. In a language of vibration, the framework imparts to the client so it can direct the user. This uses the concept of feeling for touch for visually impaired clients.

III. CONSIDERATIONS FOR SYSTEM DESIGN

For the sake of optimal performance in our system, we tried different hardware to design a system that will help these visually impaired people. The hardware, which should be selected, must be efficient in terms of battery life, night vision and weight. Similarly, we discussed and explored different techniques, but we selected the ones that were simplest and easy-to-use algorithms for implementation. Our proposed project utilizes Raspberry-Pi 4B, camera, Ultrasonic sensor and Arduino [6].

The proposed block diagram of our scheme has been shown in figure 1. Our system measures distance between the blind person and the obstacle along the path and detects objects using Viola Jones algorithm [7] and TensorFlow object detection API [8]. In first step, this system measures the distance from the obstacle using Ultrasonic sensor and Arduino for this purpose. We programmed Ultrasonic sensor and Arduino in such a way that it gives a beeping sound when there is an obstacle along its path. We used a buzzer for beeping sound. First it checks an obstacle on the front side, and if there is one, then it gives a beeping sound. After checking front side, it checks right side and then left side for obstacles, on similar grounds. At this level, it is important to differentiate the obstacles on the left and right side respectively. We handle this situation by changing the number of sounds. If there is an obstacle on the left side, it gives two beeping sounds and for right side it gives three beeping sounds. In this way, the user can differentiate between the placement of the obstacle on either side. Further, we programmed different type of sound for each side so the blind person can easily differentiate each side from the other. In addition, for a situation where the user comes across obstacles on every side, it gives four beeping sounds.

At the same time, our system is also taking pictures form the scene, and afterwards sends these pictures for pre-processing. The system pre-processes these pictures with the help of Viola Jones algorithm and TensorFlow Object Detection algorithm [9].

We used Viola Jones for face detection purpose. In this algorithm [10], the input image is scanned into Grayscale image and then features are extracted by using Haar-like feature [11]. The features are extracted by comparing pixels of different parts of face. Then sum of these pixels is calculated by using an integral image. Afterwards, we trained the Ada-boost classifier [12] to identify different features from the said image. This cascade classifier differentiates between face and non-face regions. Hence, by using the four steps of Viola Jones algorithm mentioned above, we can detect the face from the image.

In the TensorFlow object detection API, the classifier first divides the picture into a large number of small bounding boxes and then features are extracted from each bounding box. Afterwards, the overlapping boxes are bounded into a single box. In this way, objects are detected by using TensorFlow Object Detection. The presented research utilizes simple calculations for its execution. When contrasted with different frameworks, this framework is a minimal effort, convenient, reliable and simple to wear.

So, our next step is the hardware implementation of the proposed system. The proposed system uses the following components for this purpose. We have assembled the components on a stick including Raspberry pi 4 (4GB RAM), ultrasonic sensor (HC-SR04), Raspberry Pi camera (V2), servo motor (sg90) and battery as a power source. The software part of the project, on which experimental procedures are done, is performed on a computer having 8GB RAM, 4-cores processor, and a default graphic card of Intel GeForce GT 730M. Keeping these specifications in mind, our computational process doesn't seem to be lagging or causing any time delays. After training our raspberry pi, it performed even better. Further details are enlisted about working of hardware part of our project.

1) We use a Raspberry pi camera to take pictures of the scenario, whether it is an object or a person. This camera is mounted on the stick at some height so that it can capture the video as in real time. We have consulted an eye specialist and according to his recommendations we placed ultrasonic sensor at ground level as it helps detection in bumps, stairs, and hurdles in the pathway, while camera is positioned in the middle of human body. As we know that a normal lens camera captures the scene at 120 degrees from all sides. So, at this position, it will capture the images perfectly and also helps avoiding any contact with user's hands or body.
2) The servo motor is attached under the ultrasonic sensor so that it rotates in such a way that it tracks and left and right of the person in order to clear the path.

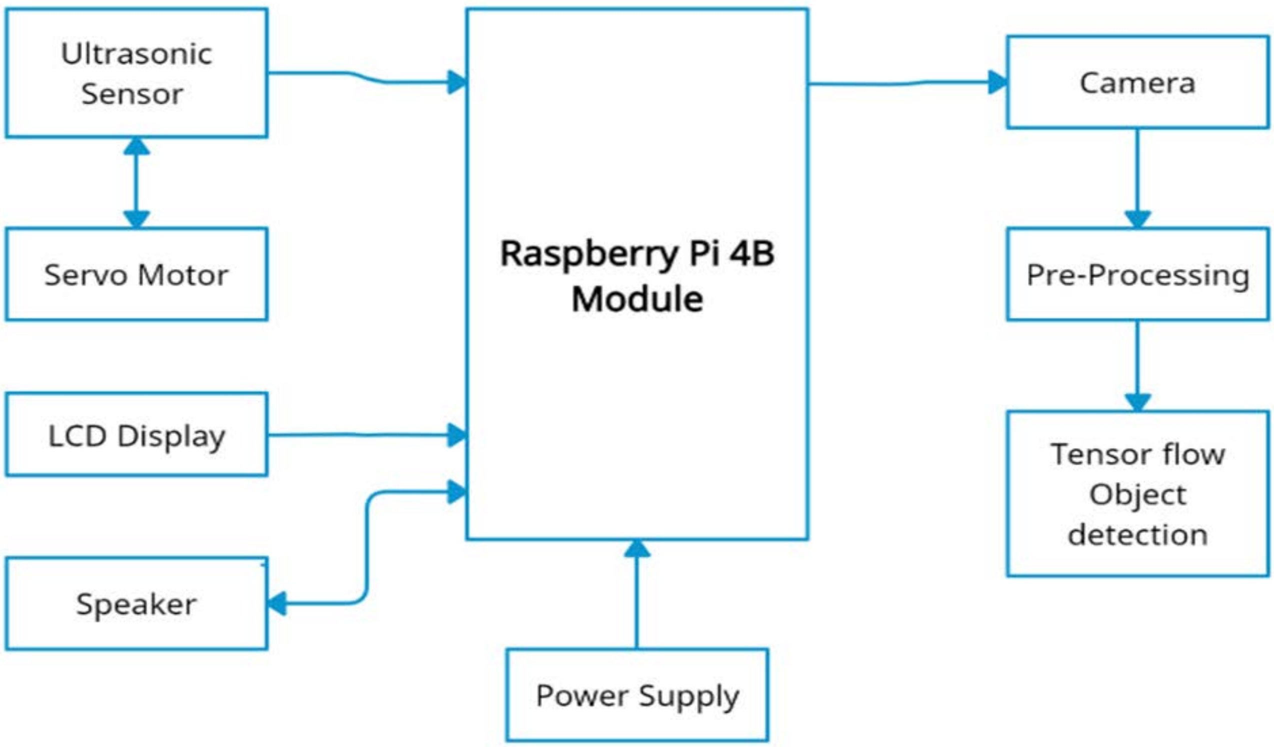


FIGURE 1. Proposed block diagram of assistive navigation system.

3) The battery is also used as a power source in order to provide the required voltage for the operation of the system which can be recharged and have impressive amount of battery timing.
4) The Raspberry pi board behaves like a focal preparing unit for observing and controlling gadgets joined to it.

A. WORKING SCHEME

The camera takes a picture, and sends it to the board along with the estimated data. The pictures taken by the camera and snags recognized by the ultrasonic sensor can be checked by the board for identifying objects around individuals as shown in Figure 2.

When the ultrasonic sensor detects any object that comes in front of the person, it will alarm the buzzer. At the same time, it will check the way on the left and right sides of the individual, and will advise the individual to turn towards left or right, where there is more space in the pathway. This ultrasonic sensor is utilized to gauge the distance of the impediments, and therefore, identifies the hindrances in the climate to give the data regarding deterrent discovery to the outwardly weakened individuals. In Figure 3, an ultrasonic sensor is mounted on a servo-motor (SG-90) [40] that rotates it at 90 degrees (clockwise and anticlockwise) towards both sides. In this way, this ultrasonic sensor detects an object from the distance of 100 cm approximately which is equal to one meter. This range can be increased or decreased depending upon the requirements but as in our system, it gives better results at the distance of one meter. The whole module is connected to battery which supplies a power of 10 Volts. We have attached our ultrasonic sensor with Arduino because our sensor is mounted on a servomotor which operates with a speed of 0.1 second per 60 degrees (0.1 s/60 degree) and its operating voltage is 4.8V. In this manner, we have used a buck-boost voltage regulator for the whole module which consists of an Arduino uno, servomotor, ultrasonic sensor,

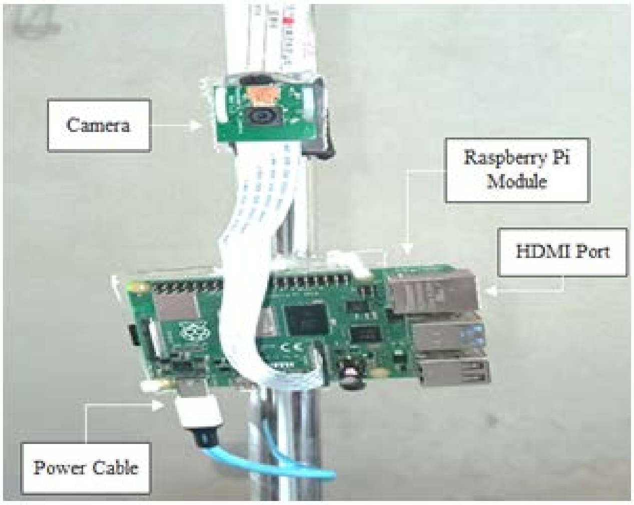


FIGURE 2. The upper part of the setup is shown. The camera is connected with Raspberry Pi module. The HDMI port connects with the computer, and electricity is provided through the power cable.

and battery of the smart cane. For further clarification, when we use servomotor on pi, its movement fluctuates a lot. However, we needed it to move exactly 90 degrees right (clockwise) and 90 degrees left (anticockwise), and then coming back to its original position if no object has been detected. On account of these technical conditions, we used Arduino separately for the sensor.

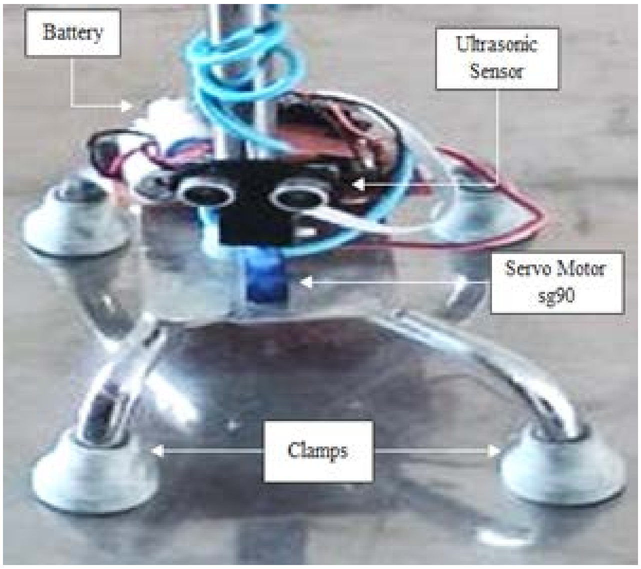


FIGURE 3. **The lower part of the setup is shown, in which the ultrasonic sensor is connected with the servomotor. For the sake of completeness, the battery and clamps have been indicated as well.**

We use a Raspberry pi camera to take pictures of the scenario, whether it is an object or a person. This camera is mounted on the stick at some height so that it can capture the video as in real time. We have consulted two eye specialists (physicians) and discussed this task with them in detail, elaborating the technical features. The discussion reveals that the visually impaired personnel are very much concerned about their movement, and the device must be developed to help them. According to their recommendations, we placed ultrasonic sensor at ground level as it helps detection in bumps, stairs, and hurdles in the pathway, thereby helping the person to check the area of movement. On the contrary, the camera being a sensitive device, must be placed at a location that is subject to least vibrations, and away from unintentional touch. Under the light of these facts, it is positioned in the middle of cane, which makes its distance apparently farthest from human body. Another fact that must be worth-mentioning is that a normal lens camera is capable to capture the scenes at 120 degrees from all sides, a fact that can be perfectly made use of when it has maximum area in the surrounding. Therefore, at this position, it will capture the images perfectly and also helps in avoiding any contact with user's hands or body. This is illustrated in figure 4.

After taking the picture, it is very important to remove noise and haze from this pictures. The reason is that these pictures have less degree of perception due to natural impacts

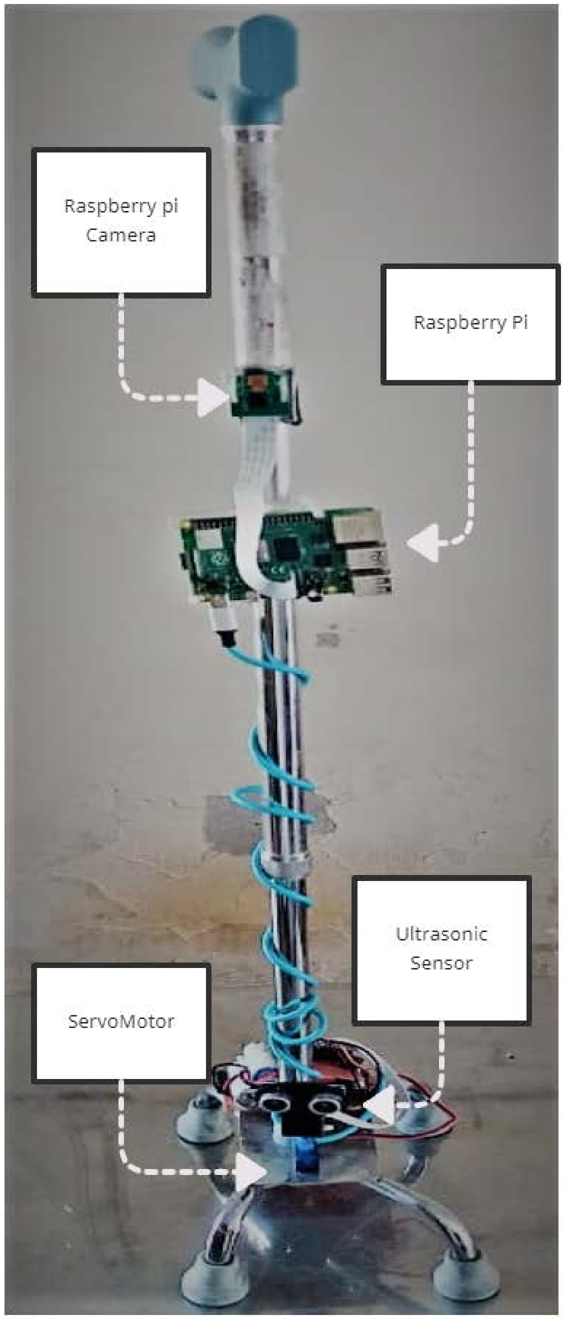


FIGURE 4. **The complete cane (sum of figures 2 and 3) is shown, with its components placed at specific locations for the handicapped, and the reasons for the placement being justified.**

on camera sensor from environment. There are several steps which are involved before the picture goes to the classification stage. First, the image is imported in the form of arrays. As some of the images which are taken from the camera differ in the size, so, we have established a standard size for all images being imported in the algorithm.

Afterwards, for noise removal, we used Gaussian Blur [43] that smoothens the images, thereby resulting in a slight reduction in the size of the images. This Gaussian filter acts as a low pass filter and removes high frequency signals. As the picture is in the form of two-dimensional array, the Gaussian function applies transformation to each pixel in the image, thereby resulting in a Gaussian distribution. In case of the two dimensions, the multiplication of both Gaussian functions [44] (one in each dimension) results in

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\left(\frac{x^2 + y^2}{2\sigma^2}\right)\right]. \tag{1}$$

Here, σ represents the standard deviation of the Gaussian distribution. As we are discussing the Gaussian function in two-dimensional plane, here x and y represent the coordinates located on X-plane and Y-plane, respectively.

The next step is the segmentation of the image in which the background is being separated from objects which helps to improve our region of interest. This process confines the image by improving edge detection and curves, thereby resulting in a better pre-processed dataset. In this way, the picture is now ready to be categorized further, explicitly undergoing the classification process.

IV. CLASSIFICATION THROUGH VIOLA JONES ALGORITHM

The Viola–Jones object detection framework [7], [33], which was introduced in 2001 by Paul Viola and Michael Jones, can be trained to identify the wide range of item types. The detection framework looks for characteristics that include the sums of picture pixels inside rectangular regions. Viola and Jones' features are typically more complicated since they all rely on more than one rectangular region. The value of each feature is equal to the summation of pixels inside plain rectangles less the pixel density within darkened rectangles. This type of rectangular feature is primitive. Furthermore, this scheme adapts to vertical and horizontal properties, but with a significantly coarser response.

Viola Jones algorithm is used with slight modification in the framework, as shown in figure 5, and provides much better results. The framework described has achieved remarkable results because of its implementation and integration in OpenCV. The cascaded classifier, which is being used, has combined the features efficiently. The framework is able to detect all the parts without any remarkable error. However, it must be pointed out at this level that it takes more time for processing them, so there is a need to reduce time delay and the load on our processor. This is important for the development of the sensor, as it would be used for the visually impaired people. To enhance our idea, we classify the features that are being selected through AdaBoost [47]. For this purpose, we selected a number of edge-type horizontal features and line features. Afterwards, we boosted them in order to get a strong classification. This is necessary because when there are numerous weak classifiers that are being used for face detection, we merge all those weak classifiers and convert them into a strong classifier. This in turn resulted in better accuracy and less time consumption.

All parts of the image can be distinguished in this way as they are not merged with each other. It gives an output that has a very good quality with appropriate attributes. This procedure for a precise recognition of upper body area (object) works in particular on the enhancement of face, and diminishes the difficult assignment of objects identifying upper body areas from unconstrained still pictures and video. The cascade object indicator utilizes the Viola-Jones face discovery calculation to recognize individual's body and identify the face object. This model recognizes the upper body area, which is characterized as the region that constitutes the head and shoulders, just as a sort of rearrangement of chest area and the face. Furthermore, the eyes are recognized as well, dependent on the speculation that they are more obscure than other parts of the face. Hence the eyes are being detected through rectangular line Haar features. The detection of eyes through little changes in the integral image becomes possible in this manner. It contains less noise and haze. Therefore, the accuracy of this model is good and has provided us very remarkable results.

A. DETECTION SCHEME

1) First, import the images into the system.
2) Apply pre-processing techniques on the images.
3) Pass the images through the Viola Jones algorithm.
4) Detect the faces using this algorithm.
5) If the facial features are detected by the algorithm in the image, then classify it as a face.
6) If the facial features are not detected, then discard the said image.
7) Check for any possible faces to detect in the image.
8) If any further facial features are found, classify each of them as an image, similar to the steps that have been provided above.
9) If there are no more facial features which could be detected by the system, terminate the program.

This algorithm is divided into four stages.

1) Haar Feature Selection
2) Creating an Integral Image
3) Adaboost Training
4) Cascading Classifiers

B. HAAR FEATURES

All human faces have certain characteristics in common. Haar Features can be used to match these regularities. Human faces have a few features, such as the eye region being darker than the upper cheeks and the nose bridge being whiter than the eyes. Numerous qualities combine to produce some matchable face features like eyes, lips, and nose, and their values are determined by directed gradients of pixel intensities. Viola and Jones utilized two-rectangle features. There are three types: two, three, and four-rectangles,

$$F(Haar) = \sum F_{White} - F_{Black}, \tag{2}$$

where F_{White} represents pixels in white area and F_{Black} represents pixels in black area.

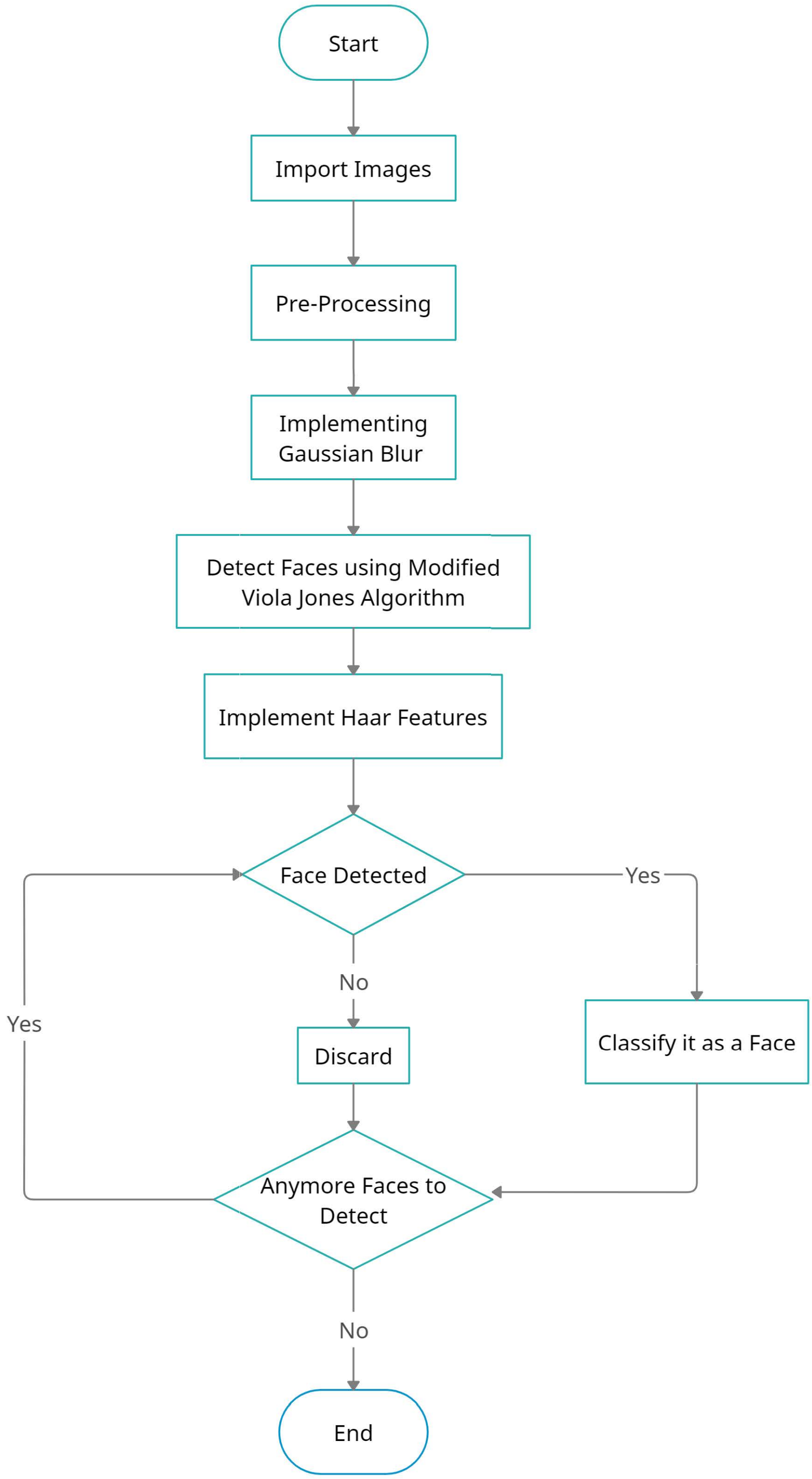


FIGURE 5. Demonstration of various steps that are involved in Viola Jones Algorithm.

C. CREATING AN INTEGRAL IMAGE

The integral image is an image representation that analyses rectangle features in real time, thereby providing them a performance edge over more comprehensive versions. Because each feature's rectangular region is always next to at least one other rectangle; any two-rectangle feature, any three-rectangle feature, and any four-rectangle feature may be derived in six array references, eight array references, and nine array references, respectively.

D. ADABOOST TRAINING

To choose the most useful characteristics and train classifiers that make use of them, the object identification framework leverages a variation of the learning method AdaBoost [47]. This approach creates a *strong* classifier by combining weighted simple *weak* classifiers in a linear way.

E. CASCADED CLASSIFIER

A Cascade Classifier is a multi-stage classifier that is capable of fast and precise detection. Each stage consists of a strong classifier that is produced by the AdaBoost Algorithm. The number of weak classifiers in a strong classifier grows as it progresses through the stages. A sequential (stage-by-stage) evaluation is performed on an input. If a classifier for a given step returns a negative test, the input is automatically removed. If the result is favorable, the input is passed on to the next stage. This multi-stage method, according to Viola and Jones [7], allows for the development of simpler classifiers, which can subsequently be used to quickly reject most negative (non-face) data while focusing on positive (facial) data [32]. In other words, we are discriminating the face from any other thing in the image that is being analyzed. This procedure is diagrammatically illustrated in figure 6.

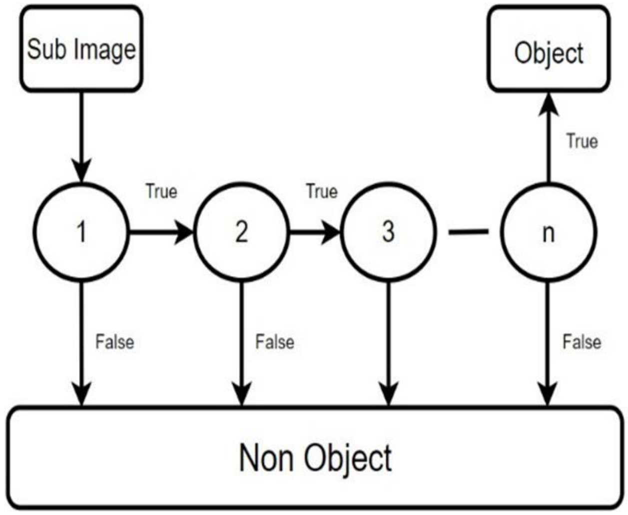


FIGURE 6. Detection of objects using cascaded classifier.

V. RESULTS AND DISCUSSION

To recognize frontal upright faces, a multilayer cascaded classifier was trained. A set of face and non-facial training pictures are used to train the classifier. The Adaboost training technique is used to train each classifier in the cascade. Similar to the discussion for Haar features, there are two types of classification of the image pixels. One is F_{white} and other is F_{Black}; a pixel in the image is compared to the gray scale and if its value is more than 0.5, it is assigned as F_{Black}, and if its value is less than 0.5, it is assigned as F_{white}. The main reason behind this is the fact that as the system works in binary values, so we classify the data as binary object whose value can either be 0 or 1.

Following the face detection experiments, different facial feature detectors are evaluated in the same way. Since they are embedded within the face, their results are shown individually to indicate that they belong to patterns of smaller size than a face or a head.

Five distinct targets are considered in the outcomes as shown in the graph in figure 8 that has been presented below; the detection of the left eye, the right eye, the two eyes together, the nose, and the mouth. A comparison is demonstrated between obtained the results and algorithm accuracies. Thus, we compared our classifications with the original Viola Jones algorithm as shown in Figure 8.

This model was utilized to assess the eye recognition and the initial attempts have shown significant outcomes. Each of the classifiers, identifies over 90% of the pictures once the face has been distinguished. Similarly, objects were also detected through trained classifiers. Using sequential evaluation, a maximum efficiency of 92% is achieved for detection.

Coco (Common objects in Context) dataset [35] is being used in our system for object detection. This is a large-scale dataset that has been used for object identification, image captioning, image segmentation and key point detection in numerous applications. In our system, we used Coco 2017 which consists of a diverse range of 123K images.

To check the performance of our system, we took a few images from the camera as to if it identifies the images correctly or not. We were satisfied and surprised to see that our system gives us quite impressive results as in the pictures shown below.

In Figure 9, our system detected a suitcase as shown in the picture with a probability of 97% (which is shown as 0.97 in the user interface of object detection). Here, the decimal in points represents the probability of the detected object which is basically derived by dividing the efficiency with hundred. Moreover, a comparison is also demonstrated to identify the difference between both of these algorithms. We can see a clear difference between figures 9a and 9b. In figure 9a, there is less opacity and contrast between foreground object and the background, in comparison with figure 9b. After the processing is done, noise has been removed from the image and hence the image becomes clearer.

Similarly, in figure 10, our system detected the keyboard which is in front of the camera with a probability of 76%. Here, our accuracy was slightly compromised because of the reason that the taken image was not precisely in the frame, so it makes difficult for the algorithm to identify it as a keyboard. After discussion among our technical team, another reason could be the difference in reflection from the various keys of the keyboard, as these might be recognized as more than a single object. This is the only drawback of the system in the sense that it identifies the object in this

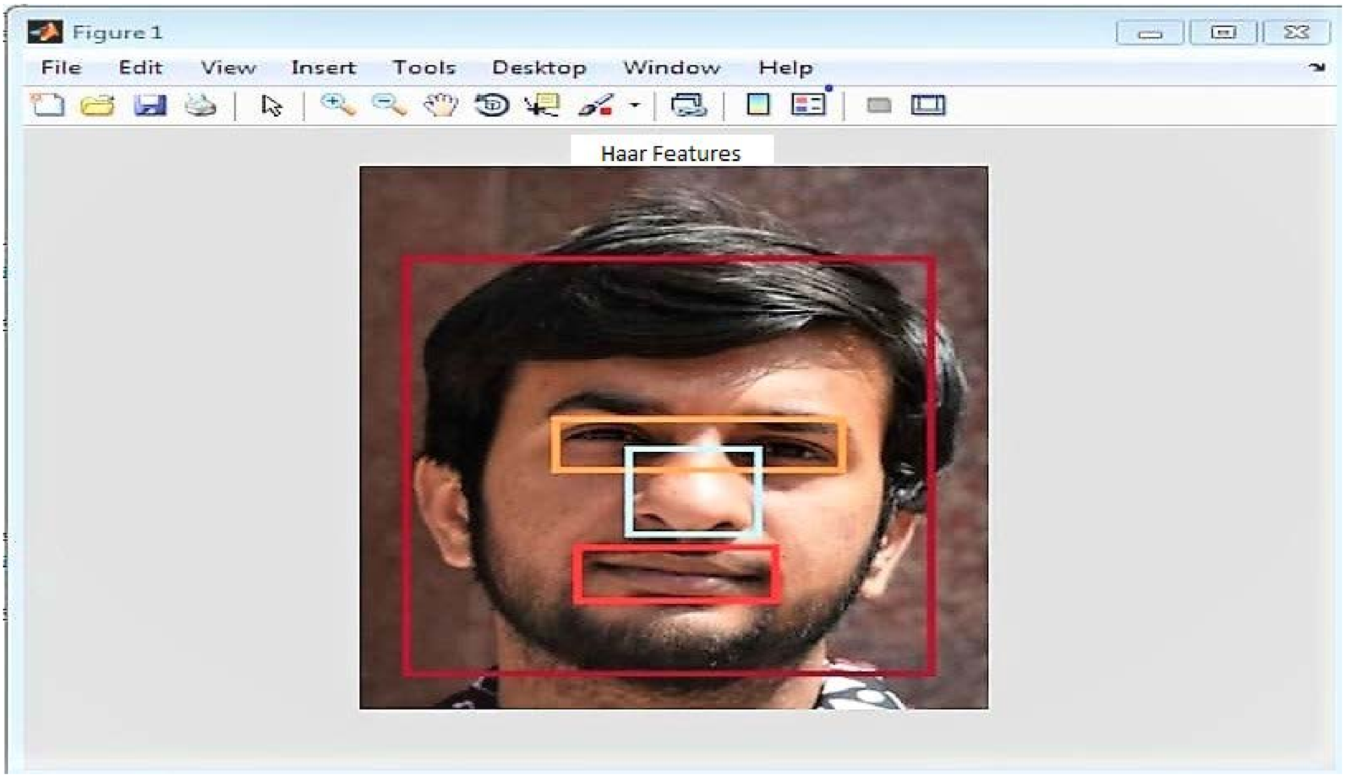


FIGURE 7. Demonstration of Haar features. The face is detected, and afterwards, the eyes, the nose and the lips are recognized.

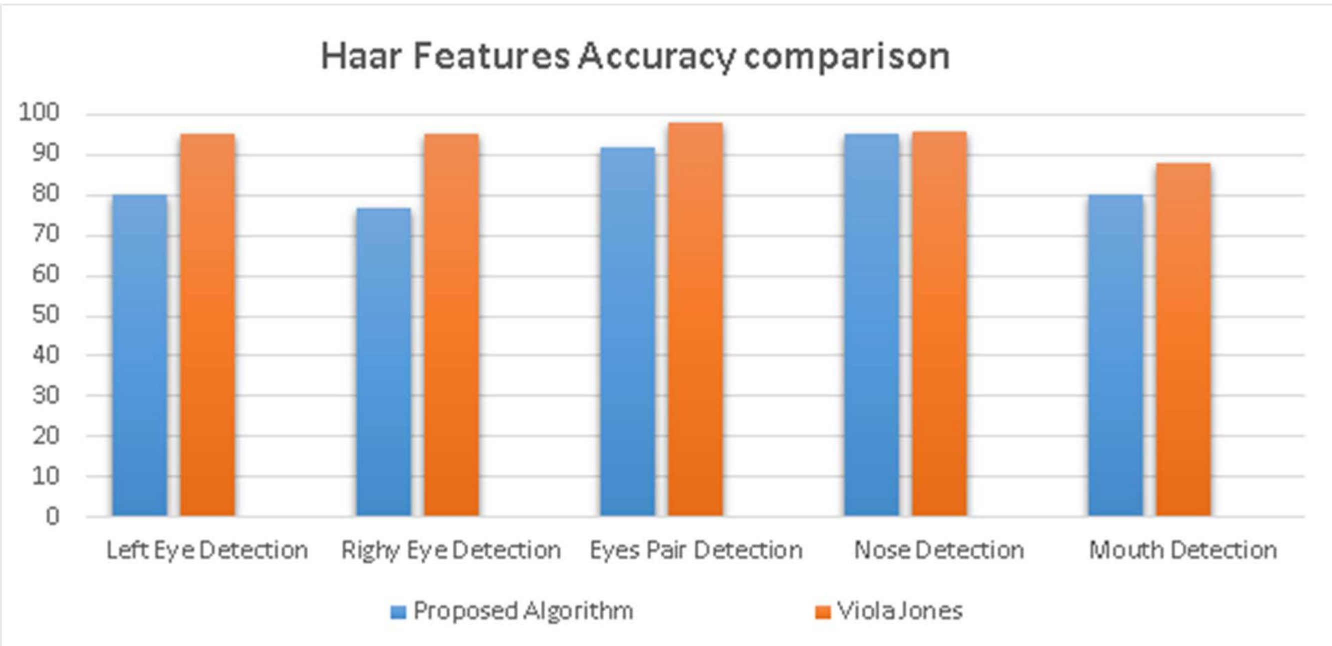
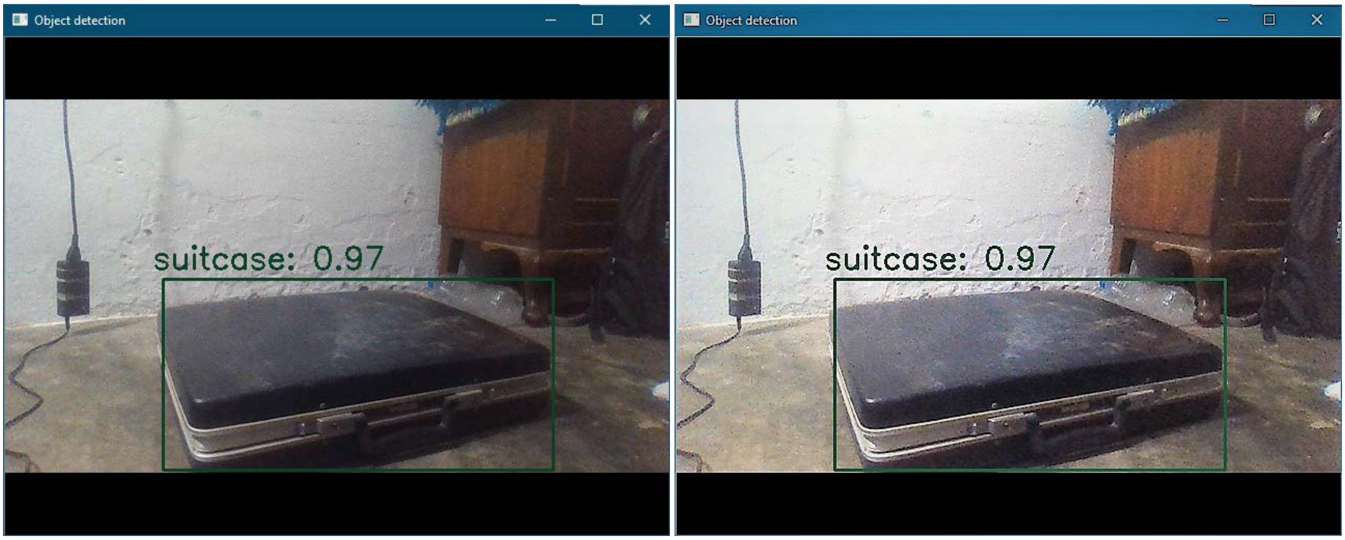


FIGURE 8. Accuracy comparison between proposed and Viola Jones algorithm.

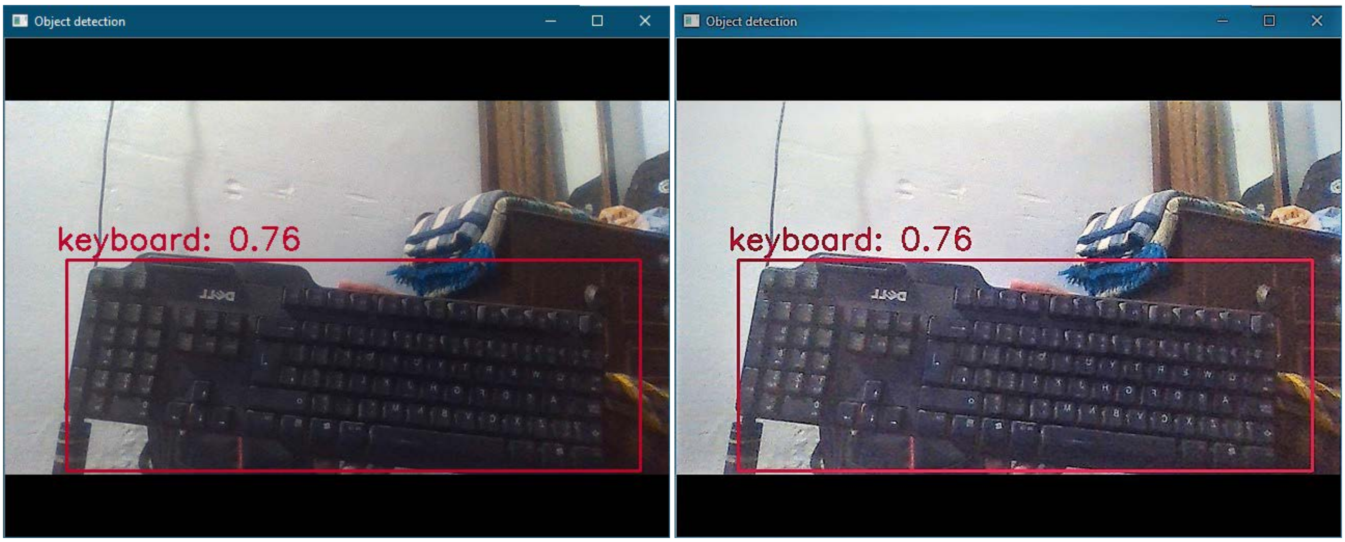
particular situation, regardless of the frame. The underlying reason is that when a person who has been visually impaired, is moving along with the stick, the camera will move as well, and shake abruptly. This effect can cause the picture to be out of frame, or distortion in the captured image. Moreover, we can see enhanced result in figure 10b which is clearer than figure 10a. However, after numerous attempts, we found that the keyboard was successfully identified, which is the main



(a) This picture is taken from original algorithm of Viola Jones and has certain amount of noise.

(b) This image is taken after a modification in Viola Jones algorithm and gives better results.

FIGURE 9. Comparison between the captured image in both cases, with a clear difference between the results.



(a) Keyboard is detected from original algorithm of Viola Jones and has certain amount of noise.

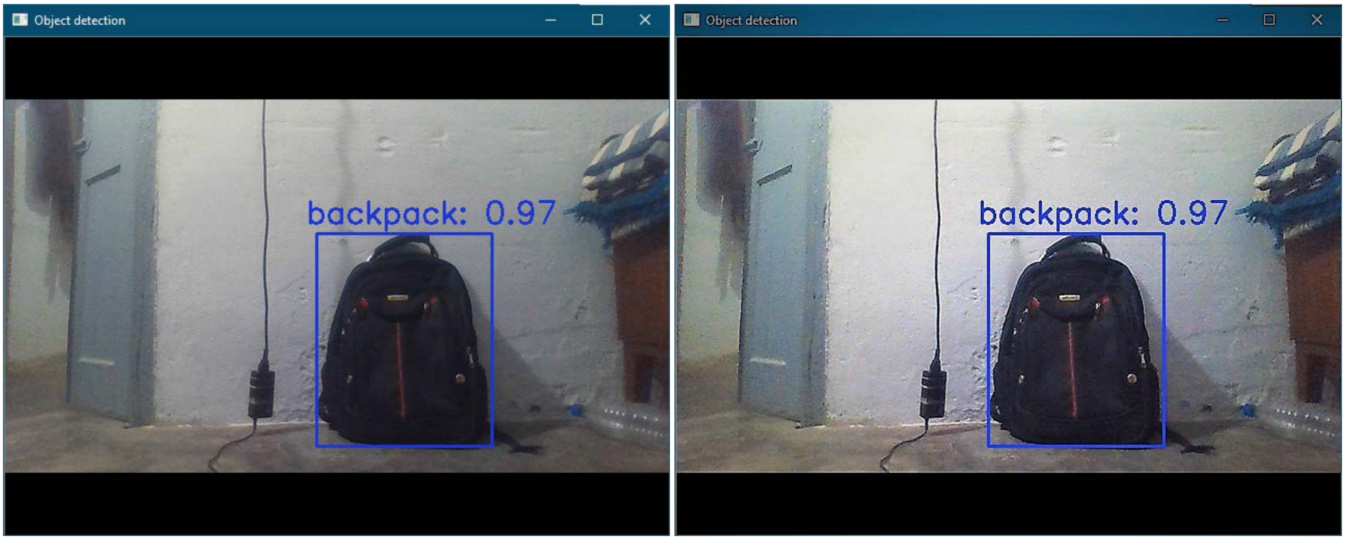
(b) Better results are gained after modification in the algorithm and contain less haze.

FIGURE 10. Comparison between the captured image in both cases, with a clear difference between the results.

objective in this situation, helping the person to recognize and avoid along the specific path.

Similarly, our system detected the image of a backpack with the probability of 0.97 which means an efficiency of 97%, as shown in the results in figure 11. Both images are compared and the second one shows better results. It has been observed that it contains much less noise and haze and has better level of brightness on foreground object. As a result, we can obviously see the backpack in fig 11b has an enhanced view than the one in 11a. This may not seem to be a worthwhile outcome at first for a normal person, however, after communicating with three visually handicapped persons, the situation becomes very important. A clear demarcation between the hindering object and the surrounding must be identified, in order for the person to pave a way to walk along in an easy way.

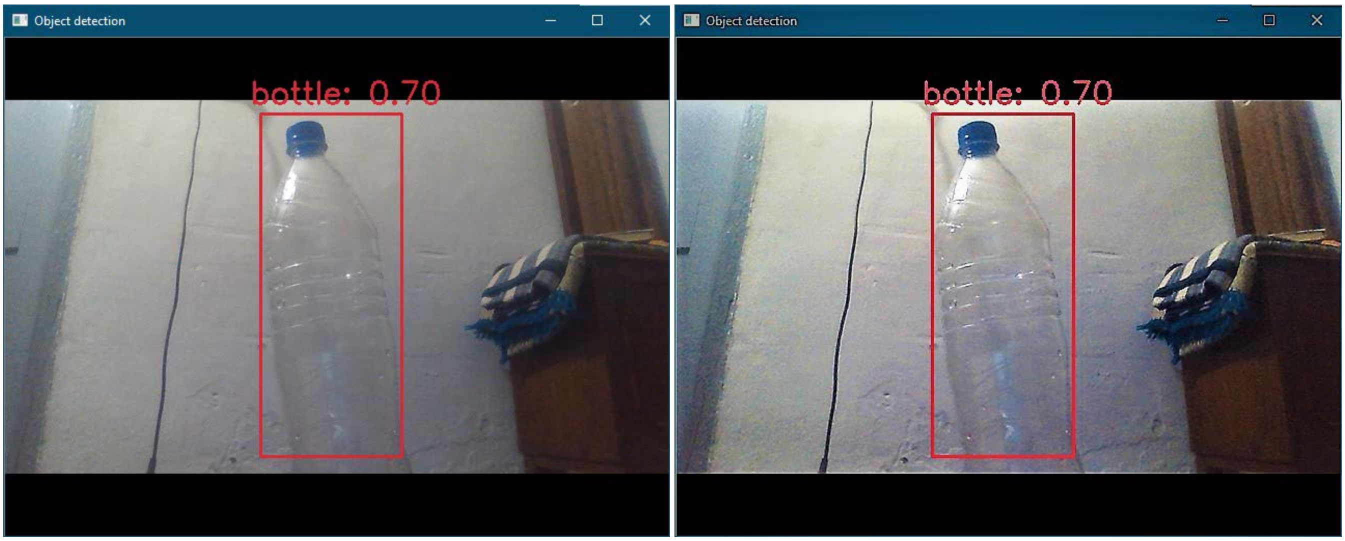
In figure 12, we have shown that our system detected a bottle from an image which has been taken from the scenario, thereby giving an efficiency of 70%. The reason behind this is obvious that the picture of bottle is not in a proper frame. The picture shows elements of colourlessness. As a result, the system finds it hard to differentiate between the background



(a) Backpack is detected from original algorithm of Viola Jones and has certain amount of noise.

(b) Better results are gained after modification in the algorithm and contains less haze.

FIGURE 11. Comparison between the captured image in both cases, with a clear difference between the results.



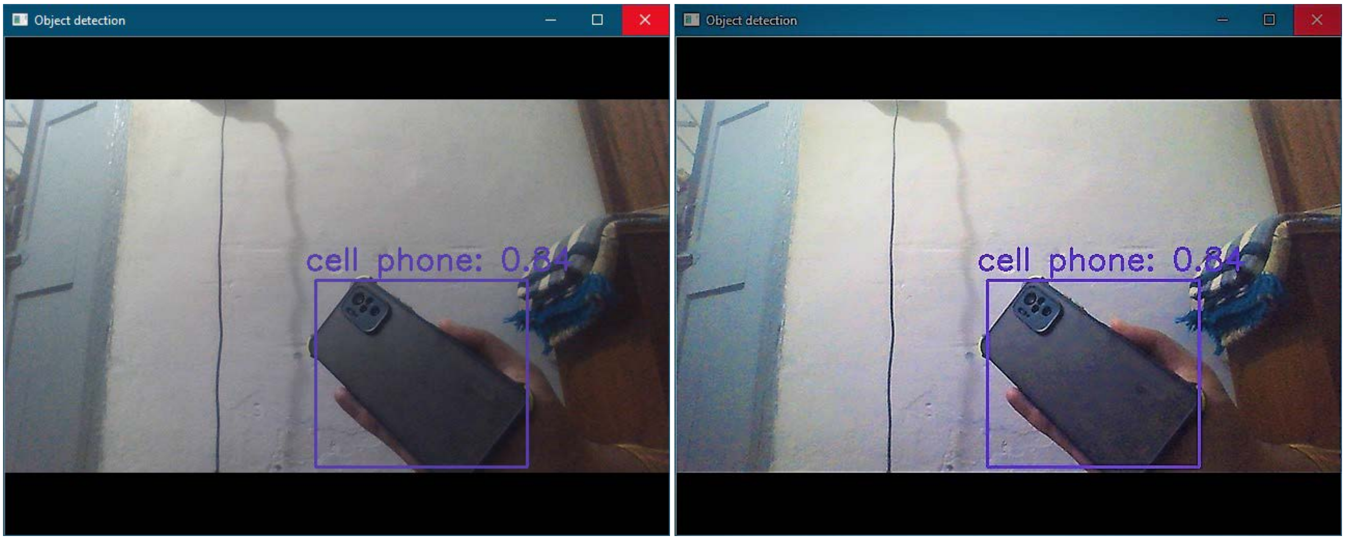
(a) Bottle is detected from original algorithm of Viola Jones and has certain amount of noise.

(b) Better results are gained after modification in the algorithm and contains less haze.

FIGURE 12. Comparison between the captured image in both cases, with a clear difference between the results.

area and the area which is covered by the bottle. To further investigate such scenarios in our system, we have tried to use another camera which is not capable of night vision, and the outcomes had been worse. After discussing with the technical staff of both cameras, we opted for the night vision camera, which lacks a bit of accuracy, but there is more brightness in the image. However, it successfully detects the periphery of the object, which, again, is the main aim for the visually handicapped personnel.

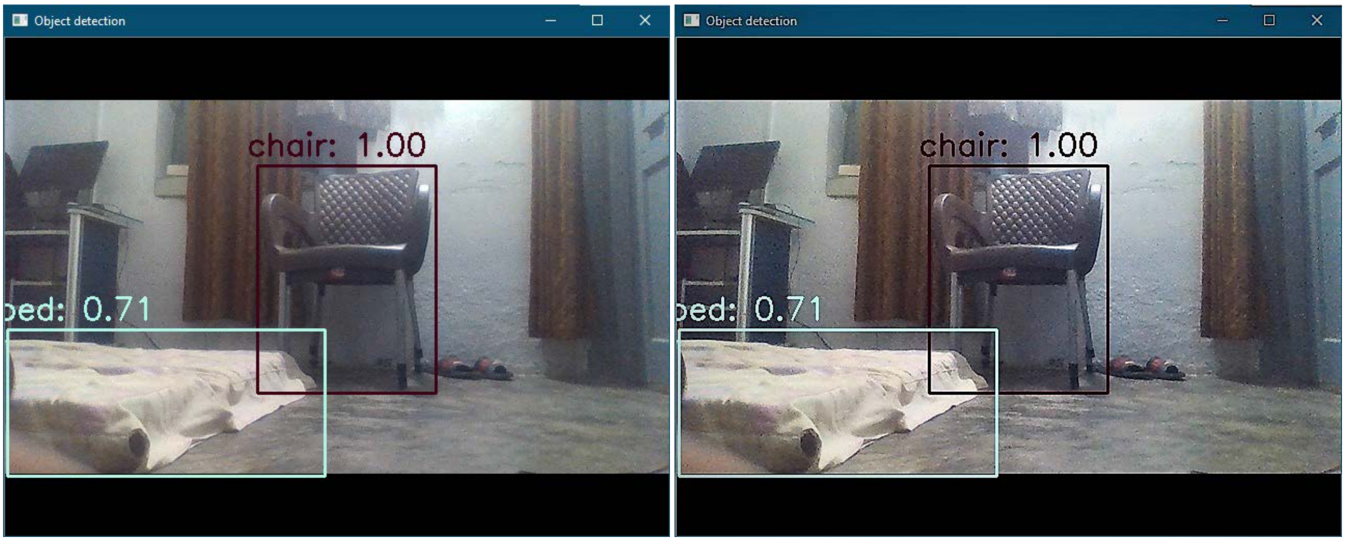
In figure 13, our system detected a cell phone in the hands of user with an efficiency of 84%. In this image, we got a better efficiency in detecting cellphones because there is a variation in the images of cellphones, on which the dataset is trained, thus better results are obtained. Moreover, we can see that figure 13a looks less bright and contains hue and moist noise in the image. On the contrary, in fig 13b, the image contains an enhanced clarity. This clearness becomes a pathway for better detection and distinction in terms of performance. In comparison to the results relating the detection of the keyboard, the efficiency is better as the fluctuation in colours does not occur on a frequent basis. We checked this by repeating this experiment four times, and the results were found to be consistently accurate.



(a) Cell Phone is detected from original algorithm of Viola Jones and has certain amount of noise.

(b) Better results are gained after modification in the algorithm and contain less haze.

FIGURE 13. Comparison between the captured image in both cases, with a clear difference between the results.



(a) Bed and Chair are detected from original algorithm of Viola Jones and has certain amount of noise.

(b) Better results are gained after modification in the algorithm and contain less haze.

FIGURE 14. Comparison between the captured image in both cases, with a clear difference between the results.

To enhance our detection scheme, we resort to check numerous obstructions in a single image. In figure 14, our system detected multiple objects as bed and chair within a single image, thereby giving a probability of 0.71 and 1.00, which means an efficiency of 71% and 100%, respectively. In this image the perspective being shown is that the system can detect multiple objects without any difficulty, as we can see that the chair can be perfectly detected. As we can see in figure 14a, there is a lack of transparency in the image as well as sharp edges for certain pixels which makes it less clear or blur. On the contrary, in figure 14b, there is no blurriness or haziness in the image, and has a clear view. In light of these facts, these are some of the bonus points of our system which will help the system to detect the obstructions in a better way, and will give a worthwhile assistance to the visually impaired user.

VI. CONCLUSION

This work is related to the design of a system for the visually impaired person that could help their lifestyle in a much better way. The system combines the functions of various components to create a multifunctional device for blind and vision impairers. The device is built in such a way that it may be used on the go. We used Viola Jones algorithm for detection purpose as the detection framework looks for characteristics that include the sums of picture pixels inside rectangular regions. Viola Jones algorithm is considered to be more complicated as more than one rectangular features is involved in the process, but it provides an ease of implementing under a confined dataset. When obstacles are identified in the path, the gadget will issue a warning through sound and haptic feedback. Because all the data is saved and processed on the Raspberry Pi, it does not require an internet access to operate. This is an extra benefit given that the internet might not be consistently available along the user's pathway. Under the scope of various circumstances that have been encountered in this work, the overall system gives us an average efficiency of about 91% which is also a great enhancement for our project. Moreover, it has a rechargeable battery whose time is around 24 hours, so the user can recharge it during night. As the system is integrated with VNC Viewer, it can be connected to cell phone of the person. The KNFB reader may be used to convert text to voice. With a single tap, the KNFB Reader transforms text into high-quality voice, providing accurate, rapid, and efficient access to single and multi-page texts. We believe that this system can be a steppingstone for greater advancements in helping visually impaired individuals.

To extend this work, we plan to add a text-to-speech system and embed it with the GSM module so that blind person can actually hear the directions in form of voice. In this way, the user can connect with the family and loved ones by sharing the precise location, through GSM. This can also be helpful if the visually impaired person loses the specific route that has to be followed. For connecting the text-to-speech system with the cell phone, the user can use a paid application like the KNFB reader [45], [46], which can be used to convert text to voice. With a single tap, the KNFB reader transforms text into high-quality voice, providing accurate, rapid, and efficient access to single and multi-page texts. We believe that this system can be a steppingstone for greater advancements in helping visually impaired individuals. This would enhance possibilities for helping not only the people with eyesight, but also paves ways to enhance the ideas of biomedical sensor for the elderly citizens [48].

ACKNOWLEDGMENT

Besides the family's moral support, the detailed discussion on a continual basis with the involved technical members of the group, as well as the vendors of the used gadgetry, is greatly appreciated.

REFERENCES

[1] H. Sharma, M. Tripathi, A. Kumar, and M. S. Gaur, "Embedded assistive stick for visually impaired persons," in *Proc. 9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2018, pp. 1–6, doi: 10.1109/ICCCNT.2018.8493707.

[2] S. K. Jarraya, W. S. Al-Shehri, and M. S. Ali, "Deep multilayer perceptron-based obstacle classification method from partial visual information: Application to the assistance of visually impaired people," *IEEE Access*, vol. 8, pp. 26612–26622, 2020, doi: 10.1109/ACCESS.2020.2970979.

[3] K. Naqvi, B. Hazela, S. Mishra, and P. Asthana, "Employing real-time object detection for visually impaired people," in *Data Analytics and Management* (Lecture Notes on Data Engineering and Communications Technologies), vol. 54, A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, Eds. Singapore: Springer, 2021, pp. 285–299, doi: 10.1007/978-981-15-8335-3_23.

[4] U. Masud, "Investigations on highly sensitive optical semiconductor laser based sensorics for medical and environmental applications," in *The Nanonose*, vol. 3862195554. Kassel, Germany: Kassel Univ. Press, 2015.

[5] N. E. Albayrak, "Object recognition using TensorFlow," in *Proc. IEEE Integr. STEM Educ. Conf. (ISEC)*, Aug. 2020, p. 1–5, doi: 10.1109/ISEC49744.2020.9397835.

[6] Q. Zhou, J. Qin, X. Xiang, Y. Tan, and N. N. Xiong, "Algorithm of helmet wearing detection based on AT-YOLO deep mode," *Comput., Mater. Continua*, vol. 69, no. 1, pp. 159–174, 2021.

[7] M. Castrillón, O. Déniz, D. Hernández, and J. Lorenzo, "A comparison of face and facial feature detectors based on the Viola–Jones general object detection framework," *Mach. Vis. Appl.*, vol. 22, no. 3, pp. 481–494, 2010.

[8] M. N. Chaudhari, M. Deshmukh, G. Ramrakhiani, and R. Parvatikar, "Face detection using Viola Jones algorithm and neural networks," in *Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697768.

[9] N. Sakic, M. Krunic, S. Stevic, and M. Dragojevic, "Camera-LIDAR object detection and distance estimation with application in collision avoidance system," in *Proc. IEEE 10th Int. Conf. Consum. Electron. (ICCE-Berlin)*, Nov. 2020, pp. 1–6, doi: 10.1109/ICCE-Berlin50680.2020.9352201.

[10] N. N. Mohd and R. Latchmanan, "An ultrasonic sensing system for assisting visually impaired person," *J. Teknol.*, vol. 78, nos. 7–4, pp. 1–5, Jul. 2016, doi: 10.11113/jt.v78.9433.

[11] *Getting the Model From TensorFlow Hub*. [Online]. Available: https://www.coursera.org/lecture/advanced-computer-vision-with-tensorflow/getting-the-model-from-tensorflow-hub-tZDzl

[12] Z. Li, F. Song, B. C. Clark, D. R. Grooms, and C. Liu, "A wearable device for indoor imminent danger detection and avoidance with region-based ground segmentation," *IEEE Access*, vol. 8, pp. 184808–184821, 2020, doi: 10.1109/ACCESS.2020.3028527.

[13] *Face Detection System Based on Viola—Jones Algorithm*, International Journal of Science and Research (IJSR), Ahmedabad, Gujarat, 2016.

[14] U. Masud and M. I. Baig, "Investigation of cavity length and mode spacing effects in dual-mode sensor," *IEEE Sensors J.*, vol. 18, no. 7, pp. 2737–2743, Apr. 2018.

[15] U. Masud, M. I. Baig, and A. Zeeshan, "Automatization analysis of the extremely sensitive laser-based dual-mode biomedical sensor," *Lasers Med. Sci.*, vol. 35, no. 7, pp. 1531–1542, Dec. 2019, doi: 10.1007/s10103-019-02945-8.

[16] 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, Dec. 2020, doi: 10.1016/j.patrec.2019.03.008.

[17] U. Pilania, A. Kaushik, Y. Vohra, and S. Jadaun, *Smart Blind Stick for Blind People* (Lecture Notes in Networks and Systems). Springer, 2021. [Online]. Available: https://books.google.com.pk/books?id=71IzEAAAQBAJ&dq=U.+Pilania,+A.+Kaushik,+Y.+Vohra,+and+S.+Jadaun,+Smart+Blind+Stick+for+Blind+People+(Lecture+Notes+in+Networks+and+Systems)&source=gbs_navlinks_s

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. 1–5.

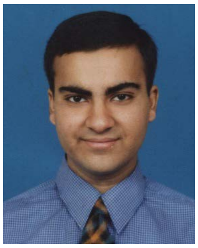
[19] M. Aghagolzadeh, H. Soltanian-Zadeh, and B. N. Araabi, "Multiscale face detection: A new approach to robust face detection," in *Proc. Int. Conf. Fuzzy Syst.*, Jul. 2006, pp. 1229–1234.

[20] K. S. Yadav and J. Singha, "Facial expression recognition using modified Viola-john's algorithm and KNN classifier," *Multimedia Tools Appl.*, vol. 79, no. 19, pp. 13089–13107, May 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11042-019-08443.

[21] I. Chawla, "Face detection & recognition using tensor flow: A review," *Int. J. Comput. Technol.*, vol. 18, pp. 7381–7388, Nov. 2018.

[22] A. Nishajith, J. Nivedha, S. S. Nair, and J. Mohammed Shaffi, "Smart capwearable visual guidance system for blind," in *Proc. Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2018, pp. 275–278.

[23] O. Butt, T. Saeed, H. Elahi, U. Masud, and U. Ghafoor, "A predictive approach to optimize a HHO generator coupled with solar PV as a standalone system," *Sustainability*, vol. 13, Oct. 2021, Art. no. 12110, doi: 10.3390/su132112110.
[24] Y. Lin, K. Wang, W. Yi, and S. Lian, "Deep learning based wearable assistive system for visually impaired people," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2549–2557, doi: 10.1109/ICCVW.2019.00312.
[25] *Embedded Assistive Stick for Visually Impaired Persons*. [Online]. Available: https://ieeexplore.ieee.org/document/8493707/
[26] P. Adarsh, P. Rathi, and M. Kumar, "YOLO V3-tiny: Object detection and recognition using one stage improved model," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 687–694.
[27] P. Banerjee, *Face Space Boundary Selection for Face Detection and Recognition*. Boca Raton, FL, USA: CRC Press, 2015, pp. 161–186.
[28] G. Hua, "Face recognition by discriminative orthogonal rank-one tensor decomposition," in *Recent Advances in Face Recognition*. 2008.
[29] R. Khan, A. Ghafoor, and N. I. Rao, "A blind image adaptive watermarking scheme for audio using wavelet transform," in *Proc. Int. Conf. Digit. Image Process.*, Mar. 2009, pp. 61–72.
[30] L. Dan, L. Chang, Z. Wei-Dong, Y. Tao, and L. Y. Hong, "Supervised non-negative tensor learning with orthogonal constraint for face detection," *J. Converg. Inf. Technol.*, vol. 8, no. 3, 2013.
[31] S. Vaidya, N. Shah, N. Shah, and R. Shankarmani, "Real-time object detection for visually challenged people," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 311–316, doi: 10.1109/ICICCS48265.2020.9121085.
[32] H. Li, "Improving the generalization capability of face spoofing detection," Ph.D. dissertation, Nanyang Technol. Univ., Singapore, 2018. [Online]. Available: https://dr.ntu.edu.sg/handle/10356/89790
[33] N. S. Ahmad, N. L. Boon, and P. Goh, "Multi-sensor obstacle detection system via model-based state-feedback control in smart cane design for the visually challenged," *IEEE Access*, vol. 6, pp. 64182–64192, 2018, doi: 10.1109/ACCESS.2018.2878423.
[34] H. Ahmad, A. Tonelli, M. Crepaldi, C. Martolini, E. Capris, and M. Gori, "Audio-visual thumble (AVT): A low-vision rehabilitation device using multisensory feedbacks," in *Proc. 42nd Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, 2020, pp. 3913–3916, doi: 10.1109/EMBC44109.2020.9175475.
[35] *Common Objects in Conext: Object Detection*. Accessed: Oct. 2,2002. [Online]. Available: https://cocodataset.org/#overview
[36] U. Masud, M. Ali, and M. Ikram, "Calibration and stability of highly sensitive fibre based laser through relative intensity noise," *Phys. Scr.*, vol. 95, Feb. 2020, Art. no. 055505, doi: 10.1088/1402-4896/ab7540.
[37] R. Joshi, S. Yadav, M. Dutta, and C. Travieso-Gonzalez, "Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people," *Entropy*, vol. 22, p. 941, Oct. 2020, doi: 10.3390/e22090941.
[38] S. Singh and B. Singh, "Intelligent walking stick for elderly and blind people," *Int. J. Eng. Res. Technol.*, vol. 9, no. 3, Mar. 2020. [Online]. Available: https://www.ijert.org/intelligent-walking-stick-for-elderly-and-blind-people
[39] Z. Xu, H. Chen, and Z. Li, "Blind image deblurring using group sparse representation," *Digit. Signal Process.*, vol. 102, Oct. 2020, Art. no. 102736, doi: 10.1016/j.dsp.2020.102736.
[40] *Servo Motor SG-90*. Accessed: Oct. 24 2018. [Online]. Available: https://components101.com/motors/servo-motor-basics-pinout-datasheet
[41] *Buck-Boost Converter: What is it*. Accessed: Oct. 5, 2020. [Online]. Available: https://www.electrical4u.com/buck-boost-converter/
[42] *Raspberry Pi 4 Your Tiny, Dual-Display, Desktop Computer*. Accessed: Aug. 15, 2020. [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
[43] M. J. Jones and P. Viola, *Fast Multi-View Face Detection*. Cambridge, MA, USA: Mitsubishi Electric Research Laboratories, 2003.
[44] M. Karimi, N. Soltanian, S. Samavi, K. Najarian, N. Karimi, and S. M. R. Soroushmehr, "Blind stereo image quality assessment inspired by brain sensory-motor fusion," *Digit. Signal Process.*, vol. 91, pp. 91–104, Aug. 2019, doi: 10.1016/j.dsp.2019.03.004.
[45] *The History of KNFB Reader*. Accessed: Aug. 20, 2021. [Online]. Available: https://knfbreader.com/about-knfb-reader
[46] L. Neat, R. Peng, S. Qin, and R. Manduchi, "Scene text access: A comparison of mobile OCR modalities for blind users," in *Proc. 24th Int. Conf. Intell. User Interface*, Mar. 2019, pp. 197–207, doi: 10.1145/3301275.3302271.
[47] Y. Freund and R. E. Schapire, "A short introduction to boosting," Tech. Rep., 1999. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.5846.
[48] U. Masud, F. Jeribi, M. Alhameed, F. Akram, A. Tahir, and M. Y. Naudhani, "Two-mode biomedical sensor build-up: Characterization of optical amplifier," *Comput., Mater. Continua*, vol. 70, no. 3, pp. 5487–5489, 2022.



USMAN MASUD received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Taxila, in 2005, the M.S. degree in electrical engineering, in Germany, in 2010, and the Ph.D. degree, in 2014. His areas of expertise include laser systems, biomedical sensors, spectroscopic applications, and wireless networks. He has been involved in multiple research areas at the moment and finds deep interest in laser-based biomedical applications. He has been an active member of Verband der Elektrotechnik, Elektronik und Informationstechnik e.V. (VDE) for several years.

TAREQ SAEED received the B.S. degree in mathematics from King Abdulaziz University (KAU), Jeddah, Saudi Arabia, the M.S. degree in financial mathematics from the University of Wollongong, and the Ph.D. degree from Griffith University, in 2018. Currently, he is working as an Assistant Professor with the Mathematics Department, KAU.

HUNIDA M. MALAIKAH received the B.Sc. degree from King Abdulaziz University, and the M.Sc. degree in applied numerical computation and the Ph.D. degree in financial mathematics from The University of Manchester, U.K. His research interests include the stock price prediction, the stock price distribution, the volatility models, numerical solution for the fractional differential equations and stochastic differential equations, and financial models.

FEZAN UL ISLAM received the bachelor's degree from the Faculty of Electrical and Electronics Engineering, University of Engineering and Technology, Taxila.

GHULAM ABBAS received the bachelor's degree from the Faculty of Electrical and Electronics Engineering, University of Engineering and Technology, Taxila.

• • •