International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI Based Support System for Blind People with an Integrated Reading Assistant

Rijwan Khan[1], Mohammad Husain[2], Jitendra*[3], Mohammad Rashid Hussain[4], Mohammad Nadeem Ahmed[5], Arshad Ali[6]

Submitted:12/03/2024 **Revised**: 27/04/2024 **Accepted**: 04/05/2024

Abstract: Mobility, navigation and reading are some of the major problems for the Blind or Visually Impaired People (BVIP). BVIP are still struggling with all of them. In the current scenario of the booming of Artificial Intelligence (AI) where everything is shifted to be "smart", we are not able to solve the problems of these BVIP. They can't even do their daily chores properly whereas we all are getting our house operated on voice assistants. As we know wearable technologies take a hike after the merging with AI. This can give the BVIP a great shift to ease their living. In this paper a prototype is presented. It is wearable like a badge on the chest or can be fitted on the spectacles that will assist the BVIP in reading documents, objects identification, and face recognition. It is implemented using the state-of-the art models provided by the Nvidia NGC. It carries out real-time video processing of visual input provided by the camera installed in it. After processing, it classifies different objects and measures their distances from it. It does this using people net and B13D proximity segmentation models of Nvidia NGC. Once the image or video analysis is done, auditory feedback is given to the user. There is a reading assistant in the gadget that can read any document present in front of it. Reading is performed using text extraction models from images.

Keywords: *Artificial Intelligence (AI), Blind or Visually Impaired People (BVIP), Nvidia NGC and Video Processing.*

1. Introduction

Blindness is one of the most common disabilities in the world. Blindness as a disability has increased in recent decades, whether due to natural causes or by accidents. People who are partially blind have cloudy vision, see only shadows, and have poor night vision or tunnel vision [1]. A person who is completely blind, on the other hand, has no vision at all. According to World Health Organization data, there are approximately 2.2 billion visually impaired or blind people worldwide. Many methods and strategies have been tested and designed to aid them and provide BVIP with some degree of comfort. Orientation and mobility is one of these methods. A trainer teaches them to walk on their own. They are taught to use their remaining senses in order to walk independently and properly. Another approach is to use guiding dogs. In this practice, dogs are carefully taught to assist blind people in moving around. The dogs go around the obstacles to warn the person to choose a different path [2]. However, BVIP may struggle to comprehend the complicated directions conveyed by these dogs. Moreover, the expense of highly trained dogs is costly. Currently, several kinds of canes, such as the white cane, the smart cane, and the laser cane, are in use [3]. However, these instruments have their own limitations. Because of its length, the canes become unhandy. Also, the cane has limits in identifying obstacles.

Many techniques have also been developed to improve the mobility of blind people. These methods are based on signal processing and sensor technologies [4]. However, these technologies usually fail in congested areas or areas with a large flow of electrical radiation. Many of the current Electronic Travel Aids (ETAs) available in the market lack real-time reading assistance and suffer from a poor user interface, high price, limited mobility, and a lack in convenience [4]. As a consequence, these gadgets are not widely used by the blind and require further refinement in design, performance, and reliability for usage in both indoor and outdoor environments.

In this paper a wearable device is proposed which is convenient to use. It will help the BVIP in doing day to day activities like walking, reading, recognizing people around them, and answering the basic queries.

Hardware components used are following:

1. Nvidia JetEx 100/300 GPU to deploy machine learning and deep learning models.

2. A camera to record the real time videos for various purposes.

[1]Prof., ABES Institute of Technology, Ghaziabad –201009, India
ORCID ID: 0000-0003-3354-3047, rijwankhan786@gmail.com
[2]Prof., Department of Computer Science, Faculty of Computer and Information Systems, Islamic University of Madinah, Kingdom of Saudi Arabia, pincode:42351, dr.husain@iu.edu.sa
[3]Assistant Prof., Galgotias University, Greater Noida –203201, India
ORCID ID: 0000-0002-5638-8601
[4]Assistant Prof., Department of Business Informatics, College of Business, King Khalid University, Abha-62217, Saudi Arabia., humohammad@kku.edu.sa
[5]Assistant Prof., Department of Computer Science, College of Computer science, King Khalid University, Abha-61421, Saudi Arabia, monaahmed@kku.edu.sa
[6]Prof. Faculty of Computer and Information Systems, Islamic University. of Madinah, Al Madinah Al Munawarah, Saudi Arabia, pincode- 42351, a.ali@iu.edu.sa
**Corresponding Author Email: jitendraofficial1990@gmail.com*

3. A microphone to take the voice input.

4. A speaker to give the command outputs or general outputs.

The camera installed in the device provides the live visual input to the system. The YOLO v7 model classifies the objects, and their distances are also measured. Regular auditory commands are given to the user so they can avoid the obstacles. It also consists of a reading assistant, which takes the image of the document as input and gives the auditory output of the written text. It also works as a full-time voice activated assistant to the user. In this paper, a prototype visual aid system for BVIP is proposed. The unique features which make our gadget different from the predefined models, includes the following: -

1. It is not dependent on the range of the sensors to guide the user about any obstacle. The distance measurement and object detection is done using the visual input, hence there is a higher level of precision and accuracy.

2. Our gadget also has the ability to interact with the user vocally, due to which users need not to find the buttons to change between different modes.

3. Real time object detection and distance measurement is performed; hence the moving objects can also be detected.

4. The device only functions on the basis of its own software and does not rely on any external sources for any signals or guidance.

We propose an AI based support system for the visually impaired or completely blind people with an additional feature of reading different documents or written texts. The system can easily be placed on a pair of spectacles of almost all types also with a mic or a headphone for the exchange of commands between user and the system the figure 1 shows the probable hardware setup of the proposed system. For the object detection YOLO v7 model is used which is basically based on pytorch framework and user algorithms like RepConv without identifying connections (RepConvN). Libraries like CvZone which basically works on the grounds of OpenCv and mediapipe libraries are used for measuring the object distance. Pytesseract which is python's optical character recognition is used for taking up textual data from the image and then python text to speech version 3 is used to convert that text in audible format. For AI assistant different libraries are used such as speech recognition, googletrans, hotword detection, openAI, dotenv which basically controls the system and interacts with the user.

In this paper, a prototype visual aid system for BVIP is proposed. The unique features which make our gadget different from the predefined models, includes the following: -

1. It is not dependent on the range of the sensors to guide the user about any obstacle. The distance measurement and object detection is done using the visual input, hence there is a higher level of precision and accuracy.

2. Our gadget also has the ability to interact with the user vocally, due to which users need not to find the buttons to change between different modes.

3. Real time object detection and distance measurement is performed; hence the moving objects can also be detected.

4. The device only functions on the basis of its own software and does not rely on any external sources for any signals or guidance.

We propose an AI based support system for the visually impaired or completely blind people with an additional feature of reading different documents or written texts. The system can easily be placed on a pair of spectacles of almost all types also with a mic or a headphone for the exchange of commands between user and the system. The figure 1 shows the probable hardware setup of the proposed system. For the object detection YOLO v7 model is used which is basically based on pytorch framework and user algorithms like RepConv without identifying connections (RepConvN). Libraries like CvZone which basically works on the grounds of OpenCv and mediapipe libraries are used for measuring the object distance. Pytesseract which is python's optical character recognition is used for taking up textual data from the image and then python text to speech version 3 is used to convert that text in audible format. For AI assistant different libraries are used such as speech recognition, googletrans, hotword detection, openAI, dotenv which basically controls the system and interacts with the user.

Fig.1. JetEx100/300 GPU

For detecting the faces of the known people with whom the user interacts on a daily basis an open-source library "Deepface" is used which basically uses VGGFace model for recognition and OpenCv for the detection of faces. Fig.1 shows the real image of the Nvidia's JetEx 100/300 GPU which is used to handle the video processing work in the system at a faster rate than any other GPU unit. Nvidia JetEx 100/300 GPU is selected as the functional device due to its great performance and configuration tweaks as well as faster model exploration, significant cost savings, and faster time to ROI as compared to the Raspberry Pi3 in the previous systems and JetEx also supports high portability.

2. Related work

It will design, implement, and validate an original non-invasive hardware and software system. This representation will be continuously constructed, updated, and given to a blind person in real time. This technology will assist people with vision impairments in any location (indoors or outdoors), without the requirement for predetermined tags or nearby sensors [5]. Imadu, et al.,[6] developed a system for visually challenged users to follow when walking. The user steers while pushing a walking stick along a sensor, which transmits the steering angle to the user by winding the handle. Zöllner, et al.,[7] showed a mobile indoor navigation system that supports BVIPs by using optical marker tracking and the Microsoft Kinect. The technology uses vibrotactile feedback to alert users to potential barriers and provide information about their surroundings. Capi and Toda [8] demonstrated a robot system to assist BVIPs in strange indoor and outdoor settings. The robotic system includes speakers, visual sensors, laser range finders, and other components that provide environmental information.

Feng et al., [9] presented a location and navigation system based on gauging wireless area network received signal strength. The system uses compressive sensing to resolve the position determination problem. The navigation module provides verbal directions as it leads the user to predefined destinations. Kammoun, et al. [10] suggested a GPS navigation system to modify Geographic Information System elements. The end users, Orientation and Mobility instructors, and user-centered design methodologies were all used. Incorporating the main classes suggested by users as well as Orientation and Mobility, a database approach is presented. The system enables better guiding of the environment used in the GIS. The system is an aid designed to improve VIPs' quality of life based on their orientation and movement abilities. Zegarra and Farcy [11] suggested uses for the inertial measurement unit (IMU) with GPS. Every second, the information is refreshed. The smartphone's user interface is created to provide heading and destination distance information.

Liao, et al. [12] suggested a VIP navigation system with RFID canes. The system does a good job of guiding traffic. The VIP can locate themselves using this system, with the help of this guided scan the RFID labels, be aware of the light signals, and discover information about the surroundings on the route. With this, the VIP can voice-direct the user's route while setting the destination. Compared to the GPS, this technology is more reliable, accurate, and cost-effective. The system gives the VIP their exact location. Nakajima and Haruyama [13] suggested an indoor navigation system for VIPs using geomagnetic field correction, LED lighting, and visible light communication technologies. The system uses visible light communication technology to offer precise location information and instant direction information. Balata, et al. [14] undertook a qualitative study to learn more about the communication problems VIPs face when navigating new environments. The experimental results suggest that VIP can cooperate with navigation and that a VIP account of the environment is sufficient for safe and reliable navigation.

Nguyen et al., [15] created a mobile robot-based Visual SLAM system to provide VIP localization services. In small or medium-sized environments, like a building or school, where GPS or Wi-Fi signals are not available, the system offers service. The system begins by gathering visual data for image acquisition. Use the Fast-Appearance Based Mapping technique to match places in a large scenario after that. To more accurately determine the robot location, the Kalman Filter is utilized. Hosny, et al. [16] designed a wheelchair-based indoor navigation system. The system is made up of three parts: a navigation system, a positioning system, and an interface for visually impaired people. The system is designed with an electric wheelchair user's best route in mind. The user selects a destination and preferred route for navigational purposes.

A wearable RGBD camera-based navigation system for VIPs was proposed by Lee and Medioni. The system consists of a smartphone interface, a real-time navigation algorithm, a haptic feedback system, and an RGBD camera mounted on glass. The mobile interface offers audio and haptic feedback as useful communication tools, while the navigation algorithm does real-time 6-DOF feature-based visual odometry using a glass-mounted RGBD camera and creates a 3D voxel model of the environment [17]. D. Ni, et al. [18] created a haptic and computer vision-based assistance walking robot system. The system is made up of a rollator structure for sturdy physical support, a Kinect for the VIP to act as their eyes and gather information about their surroundings, and ultrasonic sensors to identify symmetry roads. A wearable vibro-tactile belt is created to convey information to the VIP via vibration patterns. For the safe direction, a feature extractor method based on depth image compression is used. Bhowmick and Hazarika [19] applied information analysis and conducted a statistical analysis of the various fields of study. They have a 20-year-old comprehensive scientific research database. According to their findings, depth image compression assistive technologies for VIP are anticipated to advance over time. Nguyen, et al. [20] described a functional way-finding system designed on a mobile robot to sustain VIP. The system firstly uses outdoor technique of visual odometry to the use of indoor cover up planes through manual markers for the purpose of reliable travel in the surroundings, secondly, proposed a process to describe in optimal way the landmark of the surroundings. For the convenience of the VIP, a Smartphone interface is designed. Tapu, et al. [21] suggested a perceiving system for VIP navigation that is based on computer vision. Real-time object detection and classification algorithms based on motion are suggested. Information about the obstacle's size, nature, position, etc., is not necessary for this procedure. A method for recognizing buildings is suggested to improve the navigation and positioning capabilities.

Three subcategories of electronic assistance for the visually handicapped exist: electronic orientation aids (ETAs), positioning locating devices, and ETAs. For safe navigation, ETAs offer object detection, warning, and avoidance [22]-[24]. ETAs operate in a few simple steps: sensors are used to gather environmental data, which is then processed by a computing device to identify obstacles or other objects and provide feedback to the user. By producing a 40 kHz signal and detecting a reflected echo from the object in front of it, the ultrasonic sensors are able to detect an object within 300 cm. Based on the pulse count and time-of-flight (TOF), the distance is estimated. Ultrasonic sensors attached on smart glasses [25], [26], and boots [24] have already been suggested as a help for those who are blind. An array of infrared TOF distance sensors facing various directions is used in a novel method developed by Katzschmann et al. [27]. Villanueva and Farcy [28] use a white cane in conjunction with a near-IR LED and a photodiode to, respectively, emit and detect IR pulses reflected off barriers. It has also been done using binocular vision sensors [29], cameras [30], and cameras for the blind.

The gathered data is processed using various tools and methods. The photos obtained from the camera have been processed using a Raspberry Pi 3 Model B+ running OpenCV software [31]. There have also been instances of using Google Tango [32] platforms. Wearable technology can be used thanks to cloud computing [25]. Another choice for processing the acquired data is a field-programmable gate array [33]. To lessen noise and distortion, photos that have been collected are preprocessed. The Gaussian filter, grayscale conversion, binary image conversion, edge detection, and cropping are all manually applied to images [34]. The Tesseract optical character recognition (OCR) engine is then given the processed image in order to extract the text from it [35]. The stereo picture quality assessment [17] uses a cutting-edge method to choose the best image from a variety. A convolutional neural network (CNN), trained on huge data and operating on a cloud device, is then fed the best image. Most devices use a speaker or a headset to deliver audio feedback. A voice user interface [36] or a synthetic voice [34] produced by a text-to-speech synthesis system [37] both produce a beeping sound as the audio. In some systems, vibrations and tactile feedback are also used. And et al. [38] introduced a white cane-like haptic device with an inbuilt smart sensing approach and an active handle that detects an obstruction and generates vibrations that simulate a genuine experience on the cane handle. Guide Cane [27] is a conventional system that resembles a white cane and rolls on wheels. It features steering servo motors that guide the wheels by detecting obstructions with ultrasonic sensors. The user of this system must constantly hold the device in their hand, in contrast to the many systems that are widely available that offer a hands-free experience. The assistive gadgets NavGuide [24] and NavCane [39] employ numerous sensors to identify impediments up to knee height. Wet floor sensors are available for both NavGuide and NavCane. NavCane offers a global positioning system (GPS) with a mobile communication module that may be integrated into white cane systems. Xiao et al. [40] present a context-aware navigation framework that uses GPS to offer visual cues, distance sensing, and location-context information. With the use of Wi-Fi connectivity through the Internet, the platform may also access geographic information systems, transportation databases, and social media. A smart glass system was presented by Lan et al. [41] that can instantly detect and identify city road signs like restrooms, dining establishments, and bus stops. This method is flexible, lightweight, and portable. However, reading out the road signs by themselves might not contain enough information to make a blind user feel secure in an outdoor setting. Since public signs can vary from city to city, if a sign is not registered in the system's database, the system won't be able to identify it. Using mobile Kinect and a matrix of electrodes for obstacle detection and warning, Hoang et al. [34] created an assistive system. Because the sensors are always positioned inside the mouth while navigating, the system has a complicated setup and an uncomfortable setup. It also costs more money and is less portable.

A detailed overview of sensor-based walking aids for the blind was provided by Islam et al. in their article [42]. The authors discovered the important characteristics that make up the ideal walking assistant. Among them are designs that are affordable, straightforward, and portable and have trustworthy interior and outdoor coverage. Dakopoulos and Bourbakis [43] also established 14 structural and operational elements that describe an ideal ETA for the blind based on input from different blind user groups, software developers, and engineers. Different strategies are suggested for helping the blind and visually impaired. Cane or guide dog, infrared cane-based aid, voice-activated navigation system, laser-based walker, and ultrasonic cane are some of these methods [44]-[47]. The 2018 ICT Award has been given to this initiative. It will create and convey an audio representation of the surrounding environment to assist persons who are visually impaired. Despite extensive attempts, many existing systems frequently have cost and complexity limitations and do not fully integrate all functionalities. Our main contribution to this project was the development of a rudimentary, low-cost, portable, hands-free ETA prototype for blind people with basic indoor and outdoor text-to-speech translation capabilities. The modular design of the suggested system offers chances for future modifications and enhancements, even though it now lacks advanced capabilities like the detection of damp floors and ascending staircases, reading of road signs, usage of GPS, or mobile communication module.

3. Proposed Methodology

Fig.2 shows the complete workflow of the proposed system. The system has two main components, one is hardware which is responsible for interacting with the outside environment as well as the user. The input is taken by the camera module in the form of video footage and the microphone module takes the commands given by the user.

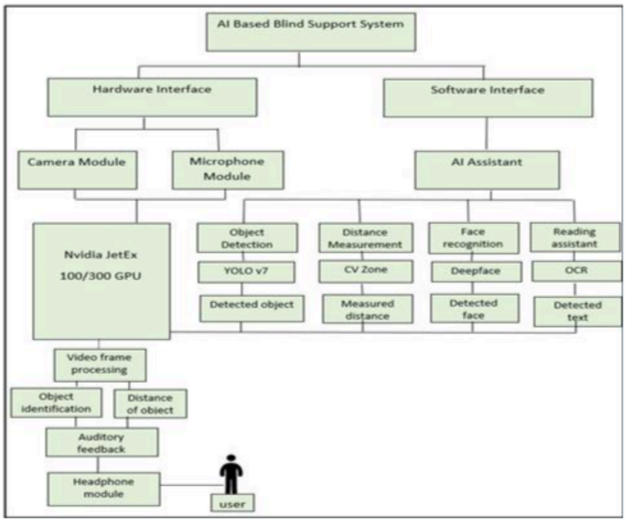


Fig.2. Complete workflow of the proposed system.

Now according to the instructions this information is processed by GPU in which video frame processing is performed to identify the object and its distance and auditory feedback is given to the user by headphone module. The other component is a software component in which the interaction with the user is done by only the A.I. assistant.

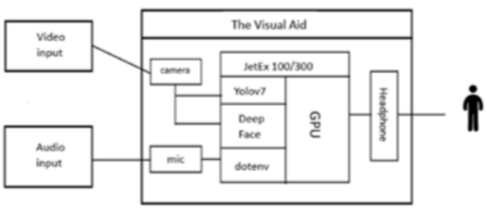


Fig.3. Hardware configuration of the proposed system

It identifies whether object detection by YOLOv7, distance measurement by CVZone, face recognition by DeepFace or reading assistant by OCR has to be performed. This information with the required module is passed to the GPU and the rest of the process is the same as mentioned above.

3.1. Object Detection

The YOLO v7 model is used for object detection which can classify the object in 80 classes. It can also be used to categorize the objects into big size objects or small size objects, accordingly warn the user about it. Relative to the size of the objects it can tell to shift a bit, or to a greater distance in order to avoid the obstacles. It can be used in both scenarios, first, in the case of an image using which the scene in front of a user can be described. Second, in the case of live feed. YOLO v7 is based on the PyTorch framework which is basically a fully featured framework for building machine learning and deep learning models, specifically those models which are used for image recognition and language processing. It is written in python and it also provides excellent support for GPUs.

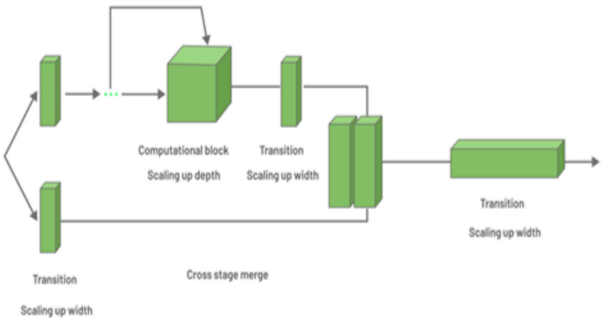


Fig.4. YOLOv7 Architecture

Fig.3 shows the hardware configuration of the proposed system. It describes that the assistant takes two types of inputs. First is video input of the surroundings using the camera module, and second is the audio commands from the user using the microphone module. The input is processed according to the instructions by the GPU through YOLOv7, DeepFace, dotenv. frameworks and output given to the user in the form of audio through the headphone module. YOLO v7 is a single stage object detection model. Here image features are extracted through a convolution neural network. Then these extracted features are combined and mixed in the neck. Finally, it predicts the locations and classes of the object as shown in Fig4. YOLO v7 is pretrained on the COCO dataset. YOLO is considered as the fastest object detection and classification model in the comparison of the other state-of-the-art models. This is achieved by extended efficient layer aggregation. In order to increase the levels of accuracy in YOLO v7, the model scaling techniques are used, it helps the model while scaling for different sizes. YOLO v7 is more robust due to the usage of gradient flow propagation paths to find which of the modules requires re-parameterization planning strategies. This model also contains an auxiliary head coarse-to-fine due to which the accuracy of the final prediction is increased by 40% as compared to other state-of-the-art models which basically adds an extra value to the precision of our system.

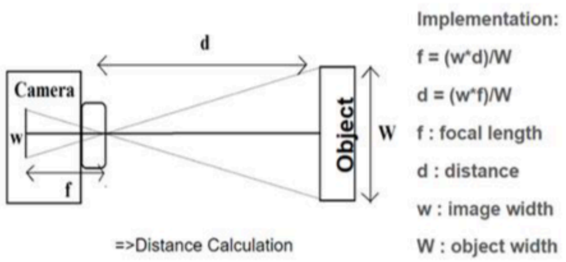
3.2. Distance Measurement

Measuring distances of distant objects is also a major task of the system. On the basis of the distance from the system or the user, the navigation commands are given to the user. When any object comes closer to the user by 50–60 inches, that object is considered an obstacle, and according to the position of that object, a command is given. Occasionally for an emergency stop when something is within 10-15 inches, then a command is given to the user, that does not react according to the previous commands.

In the proposed system, distance measurement is done by using CvZone.It is a computer vision package that makes image processing and AI functions easier to implement. It is available as an open-source library in python. It basically makes the application run simply, such as tracking, face detection, and pose estimation, which also helps in differentiating between moving objects and still objects, so moving objects can be prioritized. CvZone library in python uses opencv and mediapipe libraries at its core. Fig.5 describes a method for finding the distance of an object from the system. At first, it calculates the width of an object in the frame. Then for only one time, It calculates the focal length of the camera using the following formula:

$$F = (w * d) / W \quad \text{(i)}$$

Then we use this focal length to calculate the distance through



$$d = (W * F)/ w \quad \text{(ii)}$$

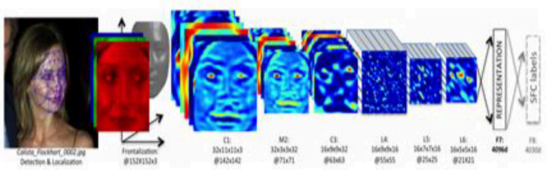


Fig.5. Explanation of the distance measurement algorithm

This algorithm calculates the distance of a person by measuring the distance between their eyes, Hence, humans can be excluded as a potential obstacle.

In the previous systems, this task of distance measurement was done using ultrasonic transducers. For example, HC-SR04, Senix, etc., whose range is fixed and also varies from sensor to sensor. Also, it cannot be placed on spectacles. Further, it is incapable of differentiating still objects and moving objects and hence more prone to accidents specially on roads. It is only limited to informing an object in front, not its exact position, to avoid it by stopping it. But our system is capable of handling these issues, hence this improves accuracy as well as the user experience.

3.3. Face Recognition

It is quite helpful for the user to recognize the face of the known person they are speaking to in order to confirm that they are not a fraudster. Face recognition is done in three phases: first, finding the face in the frame of a feed; second, detecting their features or extracting the feature points to be compared; and lastly, matching their features with the images in the database to tell who's standing in front of the user.

The face recognition in the system is done by using an open-source library, Deepface. It is basically written in python and it is a lightweight face recognition framework. It also provides the facility of facial attribute analysis such as age, gender, emotion and race. Experiments show that this model has already crossed the benchmark of 97.53% accuracy in recognizing the face. Experiments show that this model has already crossed the benchmark of 97.53% accuracy in recognizing the face.**Fig.6.**

Deepface consists of many state-of-the-art models such as VGGFace, Google facenet, openFace, facebook deepface, DeepID, ArcFace, Dlib and SFace under one library for recognition, and also have OpenCv, ResNet, ArcNet modules for the detection of face in an image. The proposed system is using Opencv for the detection of faces and VGGFace for the recognition **Fig.6.** DeepFaceArchitech(source:https://github.com/serengil/deepface).of faces. Any modern Face recognition pipeline consists of 5 common stages: detection, alignment, normalization, representation and verification. Fig.6 shows DeepFace handles all the stages in the background which increases the speed as well as the accuracy of the model. This model works on the cloud API which gives better support to any GPU on which it is executed as shown.

DeepFace is a 3D face recognition algorithm which also has some inherent advantages over 2D face recognition methods due to the large annotated 3D data. It is also a language independent package that can be used in any environment. 2D face recognition algorithms have some pitfalls due to change in lighting, different facial expressions, make-up and head orientation which will be overcome by a 3D face recognition algorithm such as DeepFace and hence this provides better accuracy in different types of environment and lighting conditions.

3.4. Reading Assistant

The majority of fraud cases against any BVIP are about taking their signature on an official document. This could be done by any imposter or by any known person concerning any legal matter. Our system also supports reading a document or any text to verify it. The system reads the textual document for the user. And hence verifies whether the conditions written on it are what have been told. This function is performed mainly in two steps: the image-to-text conversion and then speaking out that text by using different libraries such as OpenCV, PyTesseract, and Pillow for image and audio handling.

First step is to convert the image to text or to extract the text out of the image. The image is captured using the OpenCv module which is basically the easiest and fastest module for handling images or especially video frames. Then, the PyTesseract library in Python is used for recognising and reading the text embedded in the images. It is basically an optical character recognition (OCR) tool for Python. It is a wrapper for Google's Tesseract OCR engine, which is supported by the Pillow and Leptonica imaging libraries, which support all types of images such as jpeg, png, gifs, and others.

Then the extracted text is transferred to the Pytrx version 3 library of Python. It is basically a text-to-speech conversion library in Python. In this library, there is no need for an API for the language understanding of text, and hence it also works offline with the support of libraries like SAPI5, NSSS, eSpeak, etc. The text is delivered through the Tesseract OCR engine, which already supports more than 100 languages out of the box. Fig. 7 shows the workflow of the reading assistant, in which the camera puts an image as the input of a single frame and JetEx, through the OCR engine, processes it as auditory feedback.

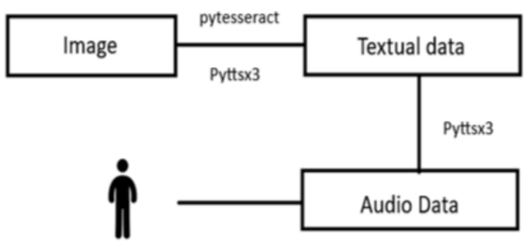


Fig. 7. Workflow of the reading assistant.

The OCR engines are more efficient in text recognition. It reduces the time of data extraction. Also, it is less prone to mistakes in the process of text extraction, hence both accuracy and security are enhanced. But it cannot recognize the handwritten documents or sometimes the letters which are overwritten on each other; in these cases, its accuracy falls.

3.5. AI Assistant

There are some major days to day chores like knowing the time, date, weather conditions, etc are some questions whose answers need to be known. Normal people can directly search on any search engine by simply typing the questions on it. For this convenience our system also has a full time AI assistant that helps in all these things and also interacts with the user and also learns about their preferences for the future use.

For this AI assistant libraries like PyAudio, openAI, dotenv, struct, math, speech recognition are used to give it different functionalities of voice recognition, interaction and speaking. In the previous systems the change of modes is done by manual buttons or switches. The user needs to remember buttons or to check it every time. With an AI assistant it can be done just by only voice commands. Hence, it gives the user freedom to control it in its own native language. The system is capable of performing autocorrections if the user says incomplete or wrong sentences. Fig.8 shows the process followed by the AI assistant. It has some defined set of commands such as switch off and switch on, same for shifting between different modes.

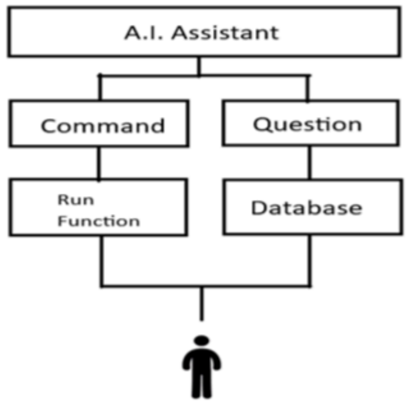


Fig.8. Workflow of the AI Assis tant.

If any command given by the user matches with these commands, then that particular function will start, otherwise it will take it as a common question and search it in logs to check if it is asked previously or not as to increase the speed of giving an answer back. If not found in the logs, then search it on the search engine using API. For the user convenience, a feature of hot word detection is used. It helps the user to start/restart the system with a clap or any other unique signal.

4. System Evaluation and Experiments

4.1. Evaluation of object identification

The YOLO v7 model is pretrained on the COCO dataset for the classification of images or objects found in an image. Model forms a bounding box on the detected objects and then tries to predict the class of the object. The dataset consists of 80 classes of images. YOLO v7 uses leaky ReLU activation function. It can detect other objects present in the background of a particular object. Hence, there would be multiple bounding boxes in a single frame, and all of them can be separated from each other very easily.

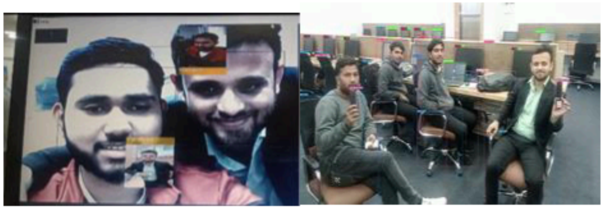


Fig.9. (a). Real time object identification (b). Real Time multiple face recognition.

Fig.9 shows the detection of multiple objects in the single frame with various confidence levels and overlapping of various objects.

Table 1: Comparison between predicted object and actual object

Test cases	Actual class	Predicted class	Failure case(s)
1	Person	Person	None
2	Monitor	Television	None
3	Monitor	Laptop (half visible)	Monitor
4	Chair	Chair	None
5	Laptop	Laptop	None
6	Remote, cellphone	Cellphone	Remote
7	Backpack	Backpack	None
8	Table	Table	None
9	Cellphone	Cellphone	None
10	Bottle	Bottle	None
11	Chair, Monitor	Chair	Monitor
12	Table, chair, TV	Chair	Television
13	Motorbike	Motorbike	None
14	Chair, person, cellphone	Chair, person, cellphone	None
15	Spoon, knife	Spoon, knife	None

On top of each bounding box, the confidence level for each predicted class is shown. Fig.9 shows the detection of multiple person's identity in a single frame. Table 1 shows the result from image as well as video object detection.

4.2. Evaluation of Distance Measurement

Table 2 shows the results of distance measured by the model of a person using coordinates of their eyes. It is using the video input from the camera. Fig.11 is showing the real time distance measurement of a person from the actual distance.

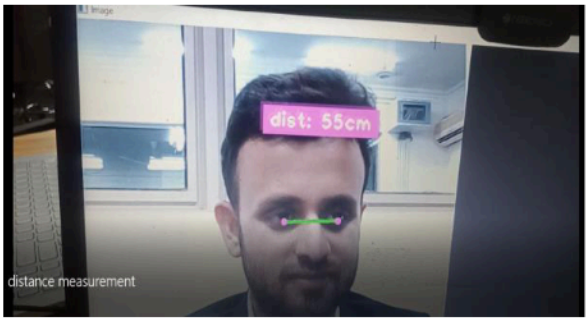


Fig.10. Detecting distance of a single object in a single frame.

Table 2: Comparison between distance recorded by camera and Ultrasonic sensor and actual distance

Test case	Text	Distance measured by Ultrasonic sensor (in cm)	Distance recorded by Camera (in cm)
1	28.6	28	28
2	75.9	75	76
3	65.6	64	65

4.3. Evaluation of Reading Assistant

The integrated reading assistant went through different types of tests. Table 3 shows the results of the tests conducted. Tests are performed under different conditions including different text styles, different colors for text, different colors of background, different lighting conditions which includes both dark and bright ambience.

Table 3. Evaluation of reading assistant in different conditions

Test case	**Text**	**Text color**	**Background color**	**Lightning conditions**	**Performance**
1	How are you	White	Black	Bright	Reads accurately
2	Legal document containing multiple lines with different fonts	Black	White	Moderate	Reads accurately but not signatures
3	"Don't judge me by my past. I don't live there anymore"	White	Blue/Black	Bright	Reads accurately

5. Conclusion

This monograph proposes an AI based helping system for the BVIP. It is wearable like a badge on the chest or can be fitted on the spectacles. It is easy to install, manage, handle, and control. The JetEx 100/300 GPU is used for better performance. Both the object detection and distance measurement is done by using visual inputs hence not relying on any sensor. It is also provided with an integrated reading assistant. It can be controlled and managed with voice commands and makes it handsfree. A detailed discussion of the software and hardware is mentioned in this paper as well as with the performance in different conditions.

6. Acknowledgements

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Research Group Project under grant number RGP.1/96/45.

7. Author contributions

Rijwan Khan[1]: Conceptualization, Methodology, **Mohammad Husain[2]:** Software, Field study, **Jitendra*[3]**: Data curation, Writing-Original draft preparation, **Mohammad Rashid Hussain[4]:** preparation, Software, **Mohammad Nadeem Ahmed[5]:** Validation., Field study, **Arshad Ali[6]:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. Xiao, S. L. Joseph, X. Zhang, B. Li, X. Li, and J. Zhang, "An assistive navigation framework for the visually impaired," IEEE Trans. HumanMach. Syst., vol. 45, no. 5, pp. 635–640, Oct. 2015.

[2] Karmel, A. Sharma, M. Pandya, and D. Garg, "IoT based assistive device for deaf, dumb and blind people," Procedia Comput. Sci., vol. 165, pp. 259–269, Nov. 2019.

[3] Y. Liu, N. R. B. Stiles, and M. Meister, "Augmented reality powers a cognitive assistant for the blind," eLife, vol. 7, Nov. 2018, Art. no. e37841.

[4] K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," IEEE Trans. Human-Mach. Syst., vol. 48, no. 2, pp. 172–182, Apr. 2018.

[5] EU H2020 Project. Sound of Vision Natural Sense of Vision Through Acoustics and Haptics. Accessed: Dec. 13, 2020.

[6] Imadu, T. Kawai, Y. Takada, and T. Tajiri, ''Walking guide interface mechanism and navigation system for the visually impaired,'' in Proc. 4th Int. Conf. Hum. Syst. Interact., HSI, May 2011, pp. 34–39.

[7] M. Zöllner, S. Huber, H.-C. Jetter, and H. Reiterer, ''NAVI—A proof-ofconcept of a mobile navigational aid for visually impaired based on the microsoft Kinect,'' in Proc. IFIP Conf. Hum.-Comput. Interact., 2011, pp. 584–587

[8] G. Capi and H. Toda, ''Development of a new robotic system for assisting visually impaired people,'' Int. J. Social Robot., vol. 4, no. S1, pp. 33–38, Nov. 2012.

[9] C. Feng, S. Valaee, A. W. S. Au, S. Reyes, S. Sorour, S. N. Markowitz, D. Gold, K. Gordon, and M. Eizenman, ''Anonymous indoor navigation system on handheld mobile devices for visually impaired,'' Int. J. Wireless Inf. Netw., vol. 19, no. 4, pp. 352–367, Dec. 2012.

[10] S. Kammoun, M. J.-M. Macé, B. Oriola, and C. Jouffrais, ''Towards a geographic information system

facilitating navigation of visually impaired users,'' in Proc. Int. Conf. Comput. Handicapped Persons, 2012, pp. 521–528.

[11] J. Zegarra and R. Farcy, ''GPS and inertial measurement unit (IMU) as a navigation system for the visually impaired,'' in Proc. Int. Conf. Comput. Handicapped Persons, 2012, pp. 29–32.

[12] C. Liao, P. Choe, T. Wu, Y. Tong, C. Dai, and Y. Liu, ''RFID-based road guiding cane system for the visually impaired,'' in Proc. Int. Conf. CrossCultural Design, 2013, pp. 86–93.

[13] M. Nakajima and S. Haruyama, ''New indoor navigation system for visually impaired people using visible light communication,'' EURASIP J. Wireless Commun. Netw., vol. 2013, no. 1, p. 37, Dec. 2013.

[14] J. Balata, J. Franc, Z. Mikovec, and P. Slavik, ''Collaborative navigation of visually impaired,'' J. Multimodal User Interface, vol. 8, no. 2, pp. 175–185, Jun. 2014.

[15] Q.-H. Nguyen, H. Vu, T. H. Tran, D. Van Hamme, P. Veelaert, W. Philips, and Q. H. Nguyen, ''A visual SLAM system on mobile robot supporting localization services to visually impaired people,'' in Proc. Eur. Conf. Comput. Vis., 2014, pp. 716–729.

[16] M. Hosny, R. Alsarrani, and A. Najjar, ''Indoor wheelchair navigation for the visually impaired,'' in Proc. Int. Conf. Hum. -Comput. Interact., 2015, pp. 411–417.

[17] Y. H. Lee and G. Medioni, ''Wearable RGBD indoor navigation system for the blind,'' in Proc. Eur. Conf. Comput. Vis., 2015, pp. 493–508.

[18] D. Ni, A. Song, L. Tian, X. Xu, and D. Chen, ''A walking assistant robotic system for the visually impaired based on computer vision and tactile perception,'' Int. J. Social Robot., vol. 7, no. 5, pp. 617–628, Nov. 2015.

[19] Bhowmick and S. M. Hazarika, ''An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends,'' J. Multimodal User Interface, vol. 11, no. 2, pp. 149–172, Jun. 2017.

[20] Q.-H. Nguyen, H. Vu, T.-H. Tran, and Q.-H. Nguyen, ''Developing a way-finding system on mobile robot assisting visually impaired people in an indoor environment,'' Multimedia Tools Appl., vol. 76, no. 2, pp. 2645–2669, Jan. 2017.

[21] R. Tapu, B. Mocanu, and T. Zaharia, ''A computer vision-based perception system for visually impaired,'' Multimedia Tools Appl., vol. 76, no. 9, pp. 11771–11807, May 2017.

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

[23] E. E. Pissaloux, R. Velazquez, and F. Maingreaud, "A new framework for cognitive mobility of visually impaired users in using tactile device," IEEE Trans. Human-Mach. Syst., vol. 47, no. 6, pp. 1040–1051, Dec. 2017.

[24] K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," IEEE Trans. Human-Mach. Syst., vol. 48, no. 2, pp. 172–182, Apr. 2018.

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

[26] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," IEEE Trans. Consum. Electron., vol. 63, no. 3, pp. 258–266, Aug. 2017.

[27] R. K. Katzschmann, B. Araki, and D. Rus, "Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 3, pp. 583–593, Mar. 2018.

[28] J. Villanueva and R. Farcy, "Optical device indicating a safe free path to blind people," IEEE Trans. Instrum. Meas., vol. 61, no. 1, pp. 170–177, Jan. 2012.

[29] B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," IEEE Internet Things J., vol. 6, no. 2, pp. 1375–1383, Apr. 2019

[30] X. Yang, S. Yuan, and Y. Tian, "Assistive clothing pattern recognition for visually impaired people," IEEE Trans. Human-Mach. Syst., vol. 44, no. 2, pp. 234–243, Apr. 2014.

[31] L. Tepelea, I. Buciu, C. Grava, I. Gavrilut, and A. Gacsadi, "A vision module for visually impaired people by using raspberry PI platform," in Proc.15th Int. Conf. Eng. Modern Electr. Syst. (EMES), Oradea, Romania, 2019, pp. 209–212.

[32] B. Li et al., "Vision-based mobile indoor assistive navigation aid for blind people," IEEE Trans. Mobile Comput., vol. 18, no. 3, pp. 702–714, Mar. 2019

[33] L. Dunai, G. Peris-Fajarnés, E. Lluna, and B. Defez, "Sensory naviga[1]tion device for blind people," J. Navig., vol. 66, no. 3, pp. 349–362, May 2013.

[34] V.-N. Hoang, T.-H. Nguyen, T.-L. Le, T.-H. Tran, T.-P. Vuong, and N. Vuillerme, "Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile kinect," Vietnam J. Comput. Sci., vol. 4, no. 2, pp. 71–83, Jul. 2016.

[35] C. I. Patel, A. Patel, and D. Patel, "Optical character recognition by open-source OCR tool Tesseract: A case study," Int. J. Comput. Appl., vol. 55, no. 10, pp. 50–56, Oct. 2012.

[36] R. Keefer, Y. Liu, and N. Bourbakis, "The development and evaluation of an eyes-free interaction

model for mobile reading devices," IEEE Trans. Human-Mach. Syst., vol. 43, no. 1, pp. 76–91, Jan. 2013.

[37] Chalamandaris, S. Karabetsos, P. Tsiakoulis, and S. Raptis, "A unit selection text-to-speech synthesis system optimized for use with screen readers," IEEE Trans. Consum. Electron., vol. 56, no. 3, pp. 1890–1897, Aug. 2010.

[38] Andò, S. Baglio, V. Marletta, and A. Valastro, "A haptic solution to assist visually impaired in mobility tasks," IEEE Trans. Human-Mach. Syst., vol. 45, no. 5, pp. 641–646, Oct. 2015.

[39] 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-Mach. Syst., vol. 49, no. 5, pp. 449–460, Oct. 2019

[40] J. Xiao, S. L. Joseph, X. Zhang, B. Li, X. Li, and J. Zhang, "An assistive navigation framework for the visually impaired," IEEE Trans. Human [1] Mach. Syst., vol. 45, no. 5, pp. 635–640, Oct. 2015.

[41] F. Lan, G. Zhai, and W. Lin, "Lightweight smart glass system with audio aid for visually impaired people," in Proc. IEEE Region 10 Conf., Macao, China, 2015, pp. 1–4.

[42] M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, "Developing walking assistants for visually impaired people: A review," IEEE Sens. J., vol. 19, no. 8, pp. 2814–2828, Apr. 2019.

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

[44] S. Al-Fahoum, H. B. Al-Hmoud, and A. A. Al-Fraihat, ''A smart infrared microcontroller-based blind guidance system,'' Act. Passive Electron. Compon., vol. 2013, pp. 1–7, Jun. 2013.

[45] A. Nada, M. A. Fakhr, and A. F. Seddik, ''Assistive infrared sensor based smart stick for blind people,'' in Proc. Sci. Inf. Conf. (SAI), Jul. 2015, pp. 1149–1154.

[46] Wachaja, P. Agarwal, M. R. Adame, K. Möller, and W. Burgard, ''A navigation aid for blind people with walking disabilities,'' in Proc. IROS Workshop Rehabil. Assistive Robot., Sep. 2014, pp. 13–14.

[47] 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,'' IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 2, pp 21, July 2011.

[48] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," IEEE Trans. Consum. Electron., vol. 63, no. 3, pp. 258–266, Aug. 2017.

[49] Blindness and vision impairment, World Health Organization, Geneva, Switzerland, Oct. 2019.

[50] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.