International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Enhancing Indoor Navigation for the Visually Impaired by Objects Recognition and Obstacle Avoidance

Mohanad A. Kadhim[1], Nidhal K. El Abbadi[2]

[1] *Computer Science department, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq*
[2] *AI Research Center, Al-Mustaqbal University, Babylon 51001, Iraq*

mohanada.araiar@student.uokufa.edu.iq, nidhal.khdhair.abass@uomus.edu.iq

Abstract: Visually impaired people face many challenges daily and depend on others a lot. They have more difficulty finding the target object, understanding it, and even navigating their internal surroundings. This paper highlights the development of a system designed to enhance the mobility and independence of blind and visually impaired individuals by recognizing their environment and navigating safely. The proposed method aims to facilitate everyday tasks by identifying objects and providing audio feedback for navigation and object detection, through a robotic body that can be worn on the head consisting of a mini-pc, a camera, and a headset. The system uses a combination of artificial intelligence algorithms, cameras, and audio feedback mechanisms to interpret visual data. YOLOv7 algorithm is used to detect indoor environments and use triangle similarity for distance estimation and alerting, with an algorithm proposed to provide directions during navigation. The F1 score of the system evaluation reveals that the system has an accuracy of 92.2%, indicating a good balance between precision and recall and showing high performance in every class. The system operates in two modes: detection and navigation, and provides audio feedback to interact with the user. This innovation seeks to address the challenges of spatial awareness and navigation for the visually impaired, with the ultimate goal of improving their quality of life and independence.

Keywords: Performance Matrices, YOLOv7 Backbone, YOLOv7 Neck, YOLOv7 Head, Triangle Similarity, Focal Length, Navigated bar.

1. Introduction

People who are blind or have visual impairments encounter obstacles in their routines because they lack visual cues, which are crucial, for most individuals. Nonetheless the human mind is incredibly versatile. Can make up for this deficiency by sharpening senses, like hearing, touch, smell and taste. Visually impaired individuals can use their heightened senses to interact with and comprehend their environment more effectively [1].

Life is a challenge for those who cannot see it. They have learnt to meet basic needs — to find their way around and move from one place to the next in their environment. While tools, like canes, can help them avoid bumping into things, they might not help them recognize things or know exactly where they are located. Consequently, people with impairments require assistance in finding and accessing things in buildings [2].

Computer Vision builds on this to enable computers to see, interpret and understand visual information within images and videos, just as well or better than humans. What it does is, it is in effect algorithm development to interpret meaningful information from visual data, as human visual perception. One of the reasons the computer vision technology helpful to the blind and visually impaired persons for better perception of the world around them [3]. Assistive technology has implemented cameras and sensors with the aid of feedback channels which increases the mobility and independence of visually impaired [4].

The main concern of the presented system in this paper is to assist persons with blindness or visual impaired to increase their attention on the surrounding objects or obstacles that would lead to impede their routine life operations. It is designed with the specific aims of identifying the internal environment and navigational to help users maneuver through different spaces and circumvent barriers, through a head strap with a camera and headset attached and a mini-PC as shown in Figure 1. The system is also designed to recognize the difference between a variety of household items, including people as well as furniture such as chairs, tables and refrigerators. Assistance in this system is provided by an audio feedback that guides the user for navigation and object recognition.

The paper is majorly divided into various sections that concentrate on a separate area of work. Literature Review section and is a comprehensive review of the methods and the research already done for the topic going to be discussed further briefly in the results section. Performance Metric: In this section, it is discussed about the performance assessment of the proposed model, showing a list of criteria on which this model is effective or not and ensuring that the results are accurate and consistent with the previous work. Distance Estimation, it goes into depth about how to estimate distances one important part of the functionality of the system.

The detailed analysis of the YOLOv7 algorithm is performed in the section "Object detection model". Introduce the Proposed System: this part gives a detailed introduction of the architecture, components, and actions in the system needed to perform the function. This section would most certainly contain lots and lots of diagrams and explanations of the system in practice, and would presumably trumpet the innovations it would make

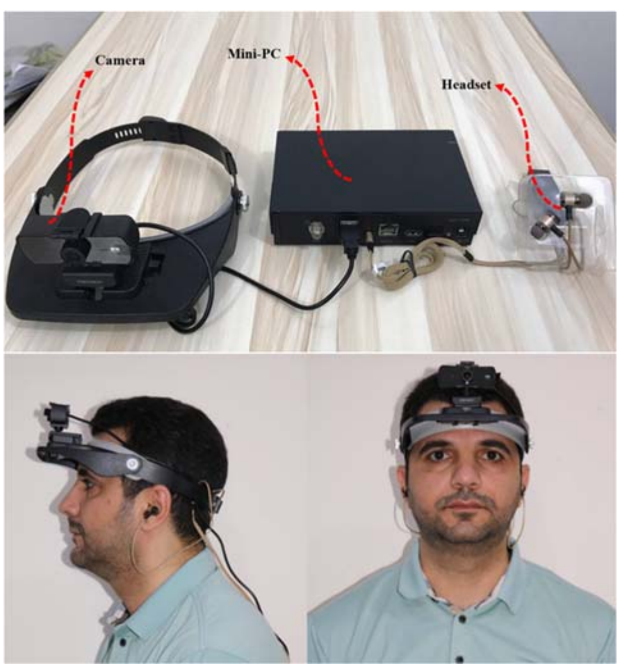

Figure 1. A head strap with a camera and headset attached to a mini-PC

in the genre it was in with design and how it was scratching the limitations of past solutions.

Experimental Results and Real Case Scenarios talks about a series of experimental results of the implemented model, and a discussion about some real-world scenarios in which the system might be used. Lastly, the last part of the paper shows the conclusion and the future work based on the proposed system. This might be new feature, optimizations, or new directions of research, that could make the system perform better or be applicable to other cases. System scalability and recommendations for further work can be planned for future developers using this section as a valuable road map.

2. Literature Survey

The study presents the development of the Android software Ayn, intended for Arabic speakers. It is meant to allow blind and visually impaired people to explain images on social media platforms. The program utilizes artificial intelligence algorithms to create written and spoken descriptions of images. Choose descriptions from the registered volunteers, or from the AI system. Users can use program to filter image descriptions according to comments or ratings [5].

The article discusses an aide system, which makes use of object recognition, combine with the Google Speech model to detect items with high accuracy and deliver the messages back audibly. The main goal of the system is to get the environment in various forms through object detection and convert it into audible speech so that people with disabilities will safely get through them. The research shows the feasibility of the proposed convolutional neural network (CNN) architecture in object detection in complex surroundings, which is considered to be the main part of helping visually impaired persons [6].

The article proposes a walking aid cane robot to detect obstacles and alert visually impaired users using one force and one ultrasound sensor. Falls can be detected by the MEMS sensor and when a fall is detected, the cane will not move at all. The notifications are sent to the caregiver and the location of the user is also sent through the cloud. There are a slew of sensors that are used to provide real-time information on the user's state and surroundings: force sensors, MEMS sensors, EEG signal sensors, ultrasonic sensors etc. The main approach of this tracking system is to minimize the accidents as well as the injuries which the incidents can make out the threat [7].

This study delves into the needs for help robots of the visually impaired in addition to discussing how it could help in living independently. More specifically, it investigates the methods to build home assistant robots, using voice command to detect and also to measure the relative position between two points inside a house from the scientific point of view. Multiple HD cameras are installed on these semi-humanoid robots to enable motion planning, autonomous movement, and object identification. Furthermore, adding to their usefulness, the robots provide the user feedback on what they have done [8].

The creation of a voice-command robotic application used to navigate blind people through interior areas is presented in this study. Users of the program can designate an indoor destination, which is thereafter linked to a map point. After that, the robot uses a traced path to lead the user to their destination. The program functioned effectively, surpassing expectations in terms of patient confidence and safety, according to usability tests with six patients. On average, it received a usability score of over eighty [9].

3. Performance Matrices

A confusion matrix is essential in measuring object detection performance, the tools used for evaluating performance are the precision, recall, and F1-score extracted from the confusion matrix. Moreover, mean average precision are used for evaluation. These metrics help in understanding how well a model can detect and locate objects within images or video frames, and they are crucial for comparing and optimizing the performance of different models. Here's a summary of the key performance metrics used in object detection [10]:

a) Precision is a measure to calculate the ratio of detected objects that were correctly identified to all detected objects. This is obtained by the division between true positives over true positives plus false positives.

$$P = \frac{TP}{TP + FP} \quad (1)$$

b) Recall is calculating the ratio of the items that have been detected correctly over all the real items is a division of True Positives by the sum of True Positives and False Negatives.

$$R = \frac{TP}{TP + FN} \quad (2)$$

c) F1 Score: use of harmonic mean of precision and recall is a fairly well balanced combination of both metrics

$$F1\,Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

d) Mean Average Precision (mAP) is a popular metric in object detection. It measures the average precision of different object categories so that it is a good indicator overall. It considers the accuracy over several recall levels and returns an average value.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (4)$$

The decision to use a single threshold, such as mAP@.5, or a range of thresholds, such as mAP@.5:.95, is dependent on the particular needs of the task that it is being used for. One threshold may suffice for some use cases that are able to tradeoff between precision and recall. Using a wide range of thresholds may be a more complete evaluation of how the model is capable on different levels of detail and variation of the data.

4. YOLOv7 Model

This is an improvement of the real time object detection model, the Yolo v7 model. This leverages the architectural improvements in previous YOLO models such as YOLOv4, scaled-YOLOv4, YOLO-R, while adding a number of useful optimizations and improvements [11] [12]. Core components of Any YOLOv7 architecture (Backbone, Neck, Head) as shown in Figure 2.

1. Backbone [13]

The core structure of the YOLOv7 focuses on feature extraction. Consists of elements

a) Convolutional Layers; These first layers apply convolutions to learn features from the input image such as edges and textures.

b) Residual Blocks; The ResNet Block class to keep deeper flow of network inspired from ResNet designs.

c) CSP (Cross Stage Partial) Connections; This connection divides the output feature map in to two part and then concatenate them through a hierarchy process to improve the flow of gradient and to reduce computational cost.

d) E-ELAN (Extended Efficient Layer Aggregation Network): This unit increases the learning capability to expand, shuffle and combine the cardinality in an efficient manner with minimal computational overhead.

2. Neck [14]

The YOLOv7 neck is responsible for absorbing features from scales while retaining the object detection capability (from small to large objects). It typically consists of

a) FPN (Feature Pyramid Network), this network combines low level and high level features to preserve spatial resolution and semantic Information.

b) PAN (Path Aggregation Network) , this network enhances the feature maps by improving the flow of information from bottom-up pathways , improving localization and detection performance.

c) SE (Squeeze and Excitation) Blocks, these blocks modulate feature maps by exciting more useful features and suppressing less useful ones.

3. Head [11]

The primary function of the YOLOv7 head is to provide the predictions by identifying bounding boxes object scores and class probabilities. This section typically consists of the following components;

a) Anchor Boxes are predefined bounding boxes of various sizes and aspect ratios are used to predict object locations more accurately.

b) Prediction Layers, these layers output the final detections, which include class probabilities and bounding box coordinates.

c) Non-Maximum Suppression (NMS) is a post-processing step that eliminates redundant bounding boxes and retains only the most confident ones.

d) YOLO Layers are Custom layers that combine multiple detection heads to predict objects at different scales.

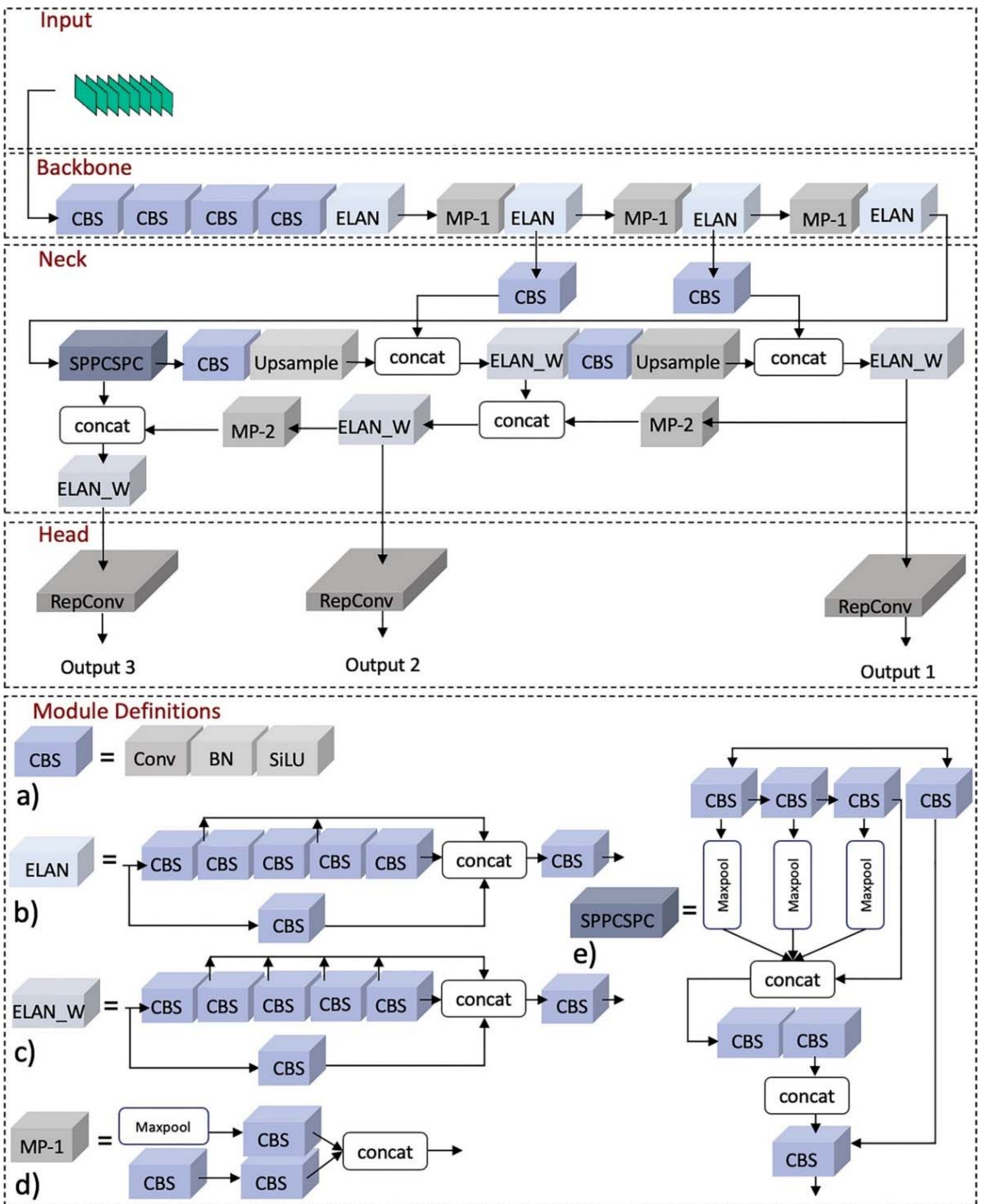

Figure 2. Architecture of YOLOv7 Model [15]

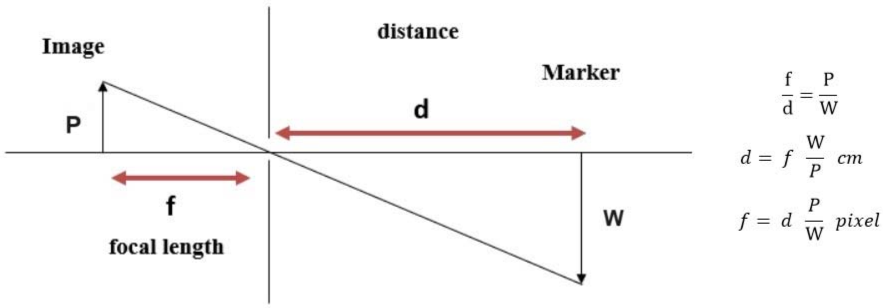

Figure 3. principle of triangle similarity, f: focal length, P : radius of the marker in the image plane, W: radius of the marker in the object plane, and d: distance from the camera to the object [18].

5. Triangle similarity

The principle of triangle similarity is used extensively in computer vision to estimate distances. This method is particularly useful when dealing with monocular vision systems where depth information is not directly available [16]. It is a technique that leverages the principles of geometry to estimate distances between objects and a camera. The core idea behind the Triangle Similarity method is that if two triangles are similar and have the same angles, their corresponding sides are in proportion [17], as we illustrate in Figure 3. Steps for Distance Calculation:

1. Measure Object Dimensions in reference image:

   A reference image is a specially captured image, in which the following values are defined (the distance between the object and the camera lens and the actual width of the object, usually measured in centimeters. In addition, the perceived width of the object in the image in pixels. These parameters are necessary to determine the focal length.

   - Perceived width of an object (PR) in pixels
   - Distance from camera to object (DR) in cm
   - Real-world width (WR) of the object.

2. Calculate Focal Length (f):

   It can be calculated by using equation (5):

$$f = D_R \frac{P_R}{W_R} \tag{5}$$

3. Detect Object in Image: Use a detection algorithm to detect the object and determine its bounding box coordinates in the image, giving perceived width (P~).

4. Calculate the distance:

   Using the principle of the similar triangle in equation (6), Then the distance equation is (7)

$$\frac{\mathrm{f}}{\mathrm{d}} = \frac{P^{\sim}}{W_R} \tag{6}$$

$$d = f \frac{W_R}{P^{\sim}} \tag{7}$$

6. Proposed system

The algorithm that has been proposed is intended to provide support for individuals who are blind or visually impaired to identify the objects in the front of camera and navigate the way indoors in real time. The proposed algorithm included three main stages: object detection and classification, object identification, and navigation as shown in Figure 4.

The input for this algorithm is a video in real time. In the first stage of the current proposal, a frame is fetched (read) from the video every 0.5 sec, where the video frames are processed as separate images. Each image is resized to 640×480 which is a recommended size for the YOLO algorithm, as shown in Figure 5.

The YOLO7 algorithm is used in this proposal, where this algorithm is trained on pre-defined images for some of the indoor objects. So, the extracted images from a video are input into the YOLO algorithm. A well-trained network will produce an efficient algorithm that enables to detection and recognition of objects accurately.

In the test phase, the YOLO is used to detect and recognize the object(s) in the input images, determine the prediction confidence for each object, and determine the coordinates (xmin, ymin, xmax, ymax) of the bounding box for each object. In this YOLO algorithm, the output is just the previous information without drawing a bounding box around the object in the image, also without annotating the objects.

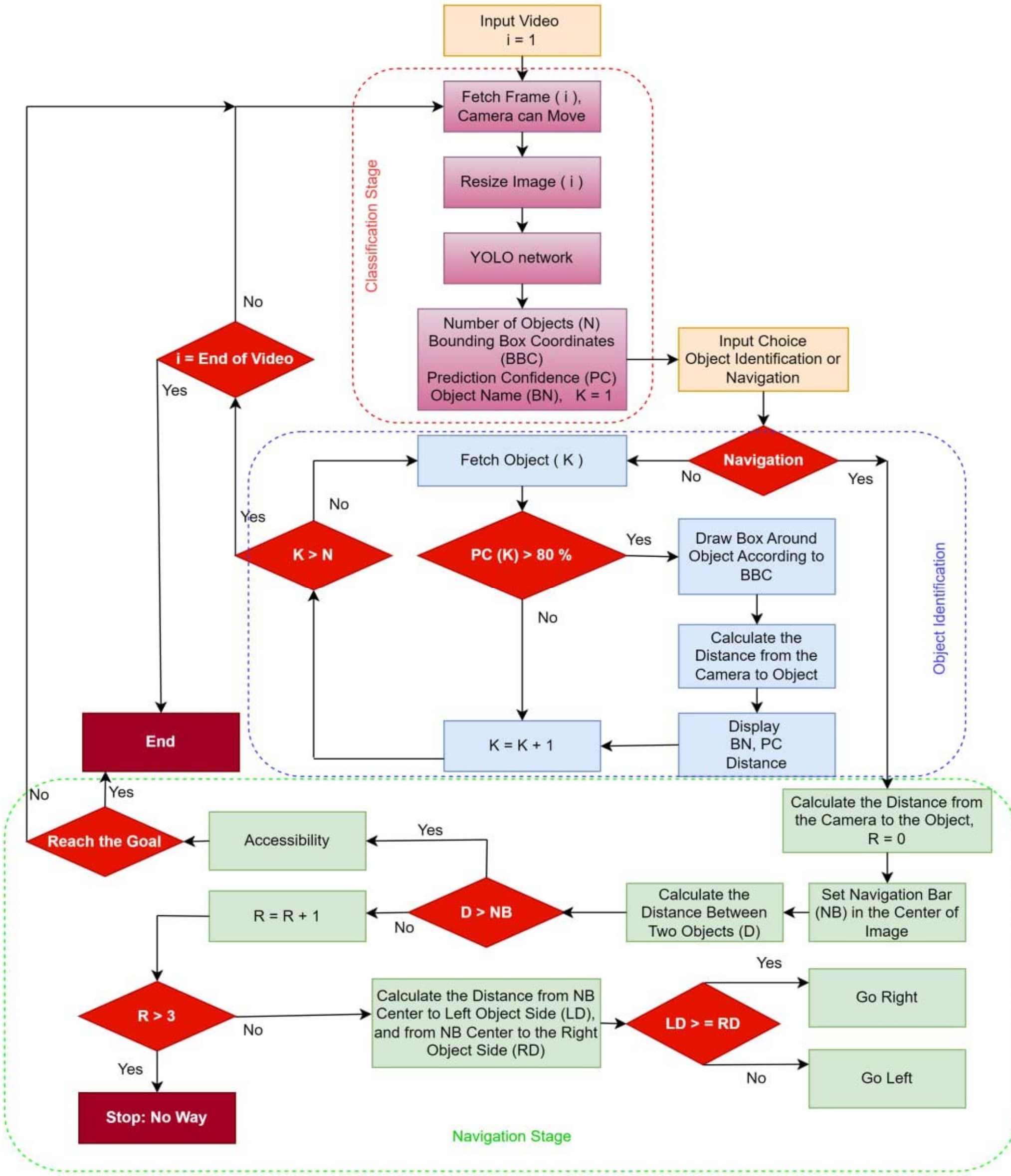

Figure 4. The block diagram of the proposed algorithm

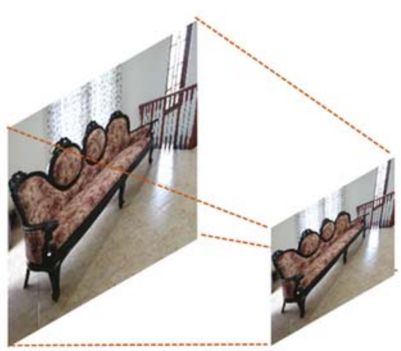

Figure 5. Resize frame image into 640×480.

The important step in this proposal after the YOLO algorithm is to provide two options for the user to select one that the proposed algorithm can employ, identify the objects in the scene or navigate the scene environment. The reason behind allowing the user to select one of two options is to enable them to either object recognition, which includes identifying objects and pronouncing their names, which allows the visually impaired to understand the scene's contents, or find the best path to their goal by overcoming obstacles' within the indoor environment, regardless the objects name or obstacles' type. However, integrating the two options might highly affect the user's concentration on finding a suitable path due to the interference between the object's name sounds and the sound of direction that guides the visually impaired to the path free of obstacles.

The second stage in this proposal is the object identification. In this stage, each object recognized in the first stage with a prediction confidence of more than 80% is selected, while the other objects are neglected. A box with dimensions determined in the first stage is drawn around the selected objects. Boxes annotated with the prediction confidence and estimated distance from the camera to the object, annotation is issued as a voice. Figure 6 shows the result of this stage.

The third stage is the navigation stage, a navigated bar (NB) is suggested to guide the impaired people in the path free of obstacles. The navigation bar as shown in Figure 7 is a circle in the center with a radius equal to 10 pixels, and two arms around the circle (one to the left and the other to the right) each with a width equal to 100 pixels. In general, the NB is centered in the center of the camera. The navigation process depends on the dimensions of the objects detected in the first stage by the YOLO algorithm. Each time the front of the camera is either an object or a path between two objects. When the NB is at the front of the path between two objects the distance between the two objects is measured (according to the dimension measured by YOLO) as shown in Figure 8, if the NB is larger than the distance measured then a voice command is issued to move forward, otherwise, the distance between the left end of the NB and right end of one of the objects at the front and also the distance between the right end of NB and the left end of other objects is measured, voice command is issued to move in the direction of less distance. The second case is the NB front of an object, in this case as previously the distance measured between the left bar end and right end of the object and right end of Nb with the left end of the object, the system issues a voice order to move in the direction of less distance. To move in the direction of less distance to the camera (generally carried by the visually impaired person) rotates at a specific angle. If he rotates three times without finding the path, then this means no path.

Figure 6. Drawing Bounding box with Class name, and prediction confidence, and distance.

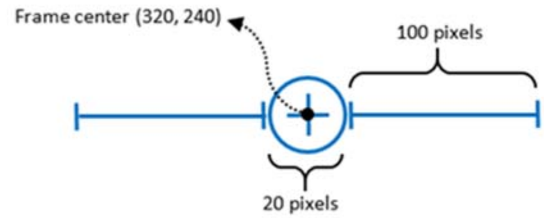

Figure 7. The shape of the Navigation Bar.

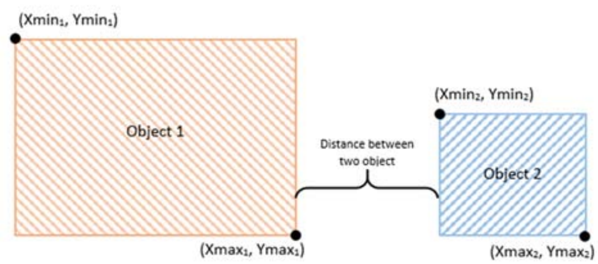

Figure 8. Distance between two objects

In the case of the robot, the suggested method rotates the robot with a specific angle toward the shorter distance or moves it toward a short distance according to the robot design. when the robot rotates three times without finding the pathway that means no path, and the robot stops navigation. In both cases, Figure 9 depicts the navigation process.

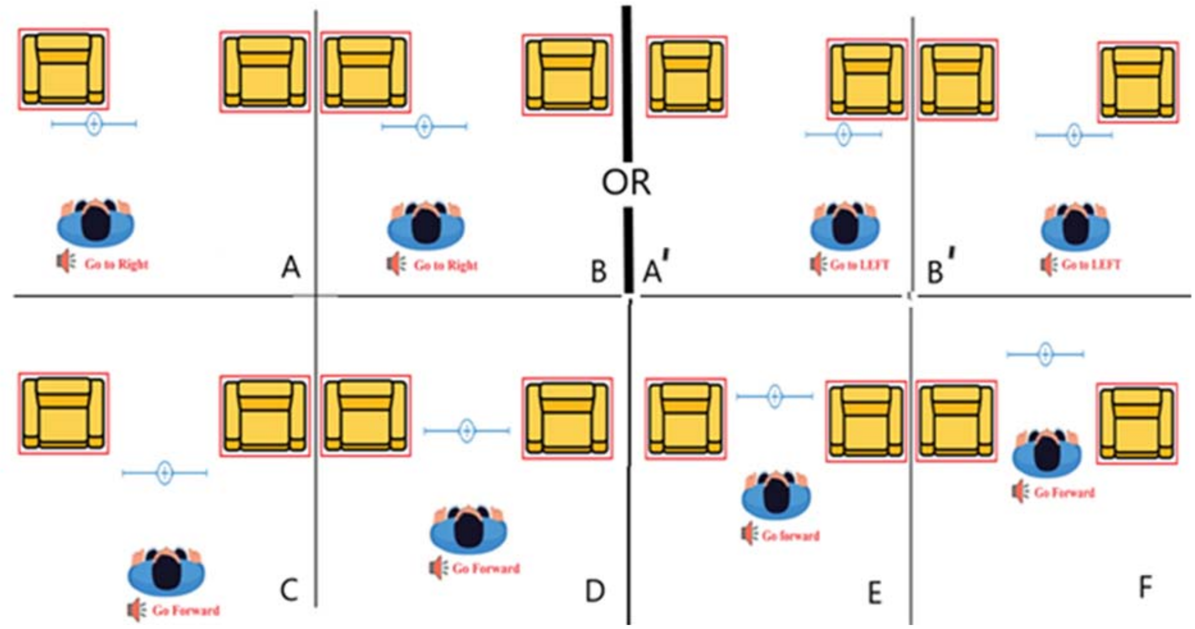

Figure 9. Explain how the system navigates through obstacles, (A and B) voice commands to move right. (A' and B') voice commands to move left. (C-F) voice commands to move forward.

7. Distance Estimation

The Triangle similarity approach, as in section 5 described, is employed to calculate the distance between the camera lens and the identified item. This approach depends on a pre-existing matrix (Ref_image_matrix) that is tailored to fit the proposed system.

The matrix comprises data obtained from reference images, each row in matrix include the camera-to-object distance, the object's actual width in centimeters, its width in pixels within the image, and an identification number assigned to each class. The array consists of reference images corresponding to the total number of classes identified by the proposed system (16 classes) as illustrate in Figure 10.

The system identifies an object in the image and retrieves the values stored in the matrix by matching the identification number of the detected object with the one in the matrix. When matching, these values are then used to calculate the focal length using Equation (5). Subsequently, the calculation of distance required figuring out of the perceived width of the detected object in the real-time stream. This information may be obtained from the bounding box coordinates provided by YOLOv7. Now, it is straightforward to determine the distance using equation (7) after all the necessary parameters are provided, including the focal length, perceived width in the image, and the object's actual width from the matrix.

```
Ref_image_matrix=[ #[0: class-id , 1: known width , 2: perceived width, 3: known distance]
                 [1,60,160,230], #human
                 [2,8,70,80], #Cell phone
                 [3,10,90,80], #bottle
                 [4,30,270,80], #book
                 [5,9,70,80], #cup
                 [6,2.5,25,80], #fork
                 [6,4,35,80], #spoon
                 [8,150,650,140], #couch
                 [9,45,220,140], #chair
                 [10,30,125,150], #handbag
                      .
                      .
                      .
                      .
                      .
                 [16,55,190,180], #refrigerator
                ]
```

Figure 10. Explain the content of reference image matrix

TABLE I. PERFORMANCE OF THE PROPOSED MODEL.

Class	*No. images*	*No. Labels*	*Precision*	*Recall*	*mAP@.5*	*mAP@.5:.95*	*F1 Score*
All	5000	5023	94.6%	90.0%	95.4%	76.3%	92.2%
Couch	5000	349	96.1%	84.2%	92.9%	77.3%	89.7%
Chair	5000	313	91.8%	89.3%	95.9%	76.7%	90.5%
Table	5000	200	95.9%	96.5%	97.4%	84.8%	96.1%
Person	5000	369	91.5%	94.3%	93.8%	89.1%	92.8%
Bottle	5000	374	93.8%	91.0%	95.9%	84.3%	92.3%
Cup	5000	338	89.3%	86.2%	91.5%	85.3%	87.7%
Fork	5000	388	90.5%	89.3%	95.6%	84.7%	89.8%
Knife	5000	394	90.0%	93.5%	89.3%	92.8%	91.7%
Spoon	5000	419	91.7%	83.3%	94.6%	80.9%	87.2%
Bowl	5000	300	85.9%	89.7%	93.8%	81.9%	87.7%
Bed	5000	369	94.3%	94.2%	89.7%	81.3%	94.2%
TV	5000	350	91.2%	88.3%	91.5%	89.4%	89.7%
Cell Phone	5000	370	95.6%	93.4%	95.3%	78.9%	94.4%
Refrigerator	5000	490	97.1%	96.1%	96.1%	78.8%	96.5%

8. RESULTS

The dataset has been carefully assembled and categorized from many datasets intended to facilitate object identification. It is structured around fourteen distinct classes, each representing a unique category of objects. 25000 images make up the dataset, of which 80% are used for training and 20% are used for testing, ensuring a balanced distribution that facilitates both learning and validation. The model has been trained using the YOLOv7 pre-trained model through the use of a deep learning method over 190 epochs.

The model's performance is evaluated across different classes of objects, including "all" (overall performance), "Chair", "Couch", "Table" and "Person", etc. For the "All" category, the model has

- The overall precision is high (94.6%), indicating that the model makes accurate positive predictions.
- Recall (90%) suggests that the model captures most of the true positive instances.
- The mAP (mean Average Precision) at different IoU thresholds (0.5 and 0.95) provides insights into the model's performance across different confidence thresholds.
- The F1 Score (92.2%) balances precision and recall.

The model is effective in detecting these objects with high accuracy, as shown in Table 1. Numerous tests have been carried out on the system to guarantee its reliability and seamless performance, ultimately enhancing the daily experiences of individuals with visual impairments. Figure 11, clearly illustrates the system's ability to accurately detect all trained objects, providing precise audio descriptions for items positioned in the center of the camera. Additionally, Figure 12, showcases the effectiveness of the proposed algorithm in navigating obstacles during movement from one location to another in two different environments. Overall, the system demonstrates impressive capabilities in aiding visually impaired individuals in their day-to-day activities.

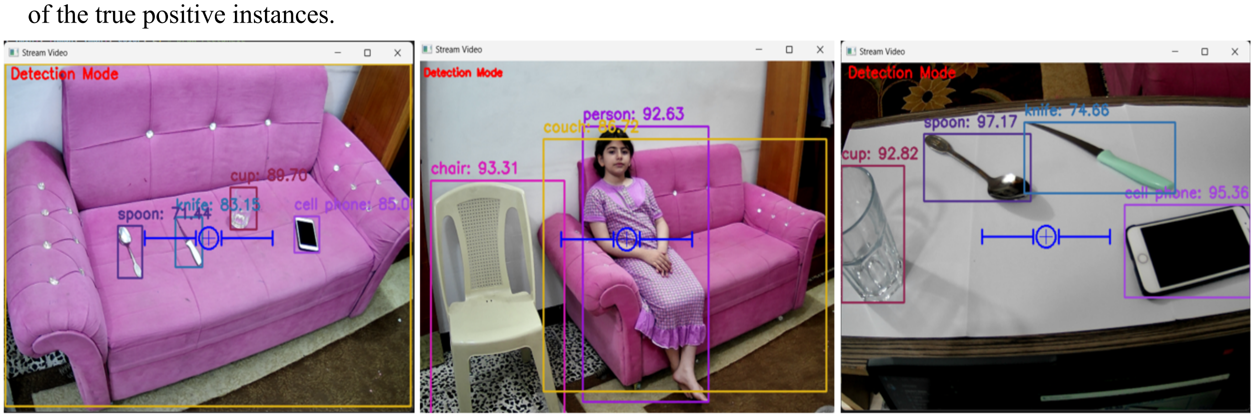

Figure 11. Detecting objects “Chair”, “Couch”, “person”, “Spoon”, “Knife”, “Cell phone”, and “cup” in multiple scenes.

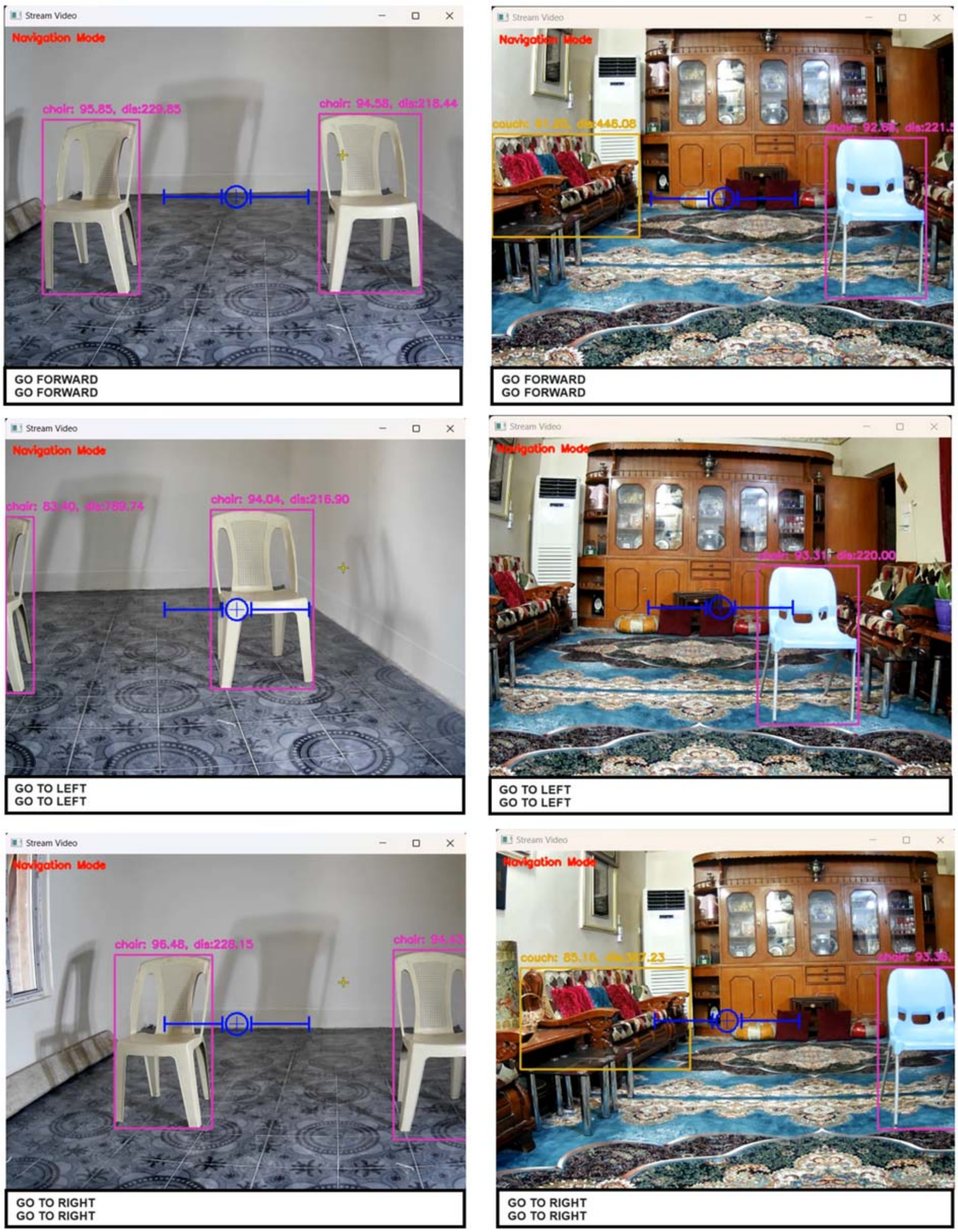

Figure 12. Testing Navigation mode in different environments.

A test was conducted to determine the best number of epochs that can give the optimum accuracy, various performance metrics were tested such as precision, recall, mAP_0.5, and mAP_0.5:0.95 as shown in Figure 13, the best number of epochs in this test was 190. Also, another test to determine the best number of epochs that gives the best F1 score is shown in Figure 14. This test confirmed that 190 epochs are optimal.

Table 2, presents a comparative analysis of existing methods that employed object detection methods to assist individuals with visual impairments. The table displays the employed methodology along with the corresponding outcomes, the level of precision achieved, and limitation.

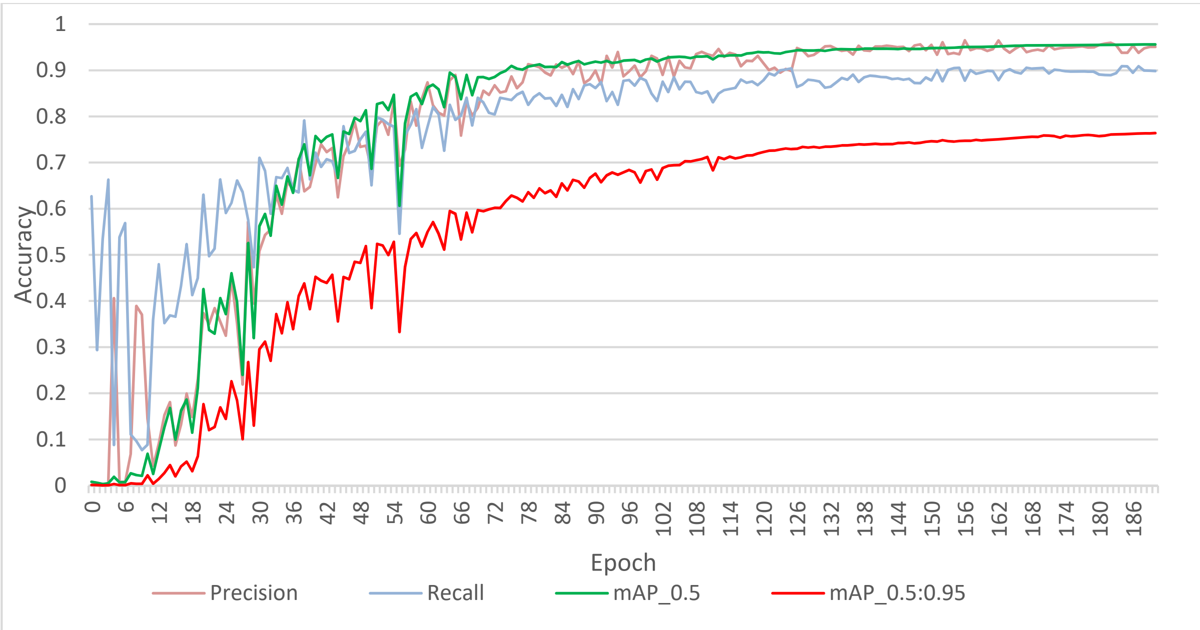

Figure 13. Performance matrices chart

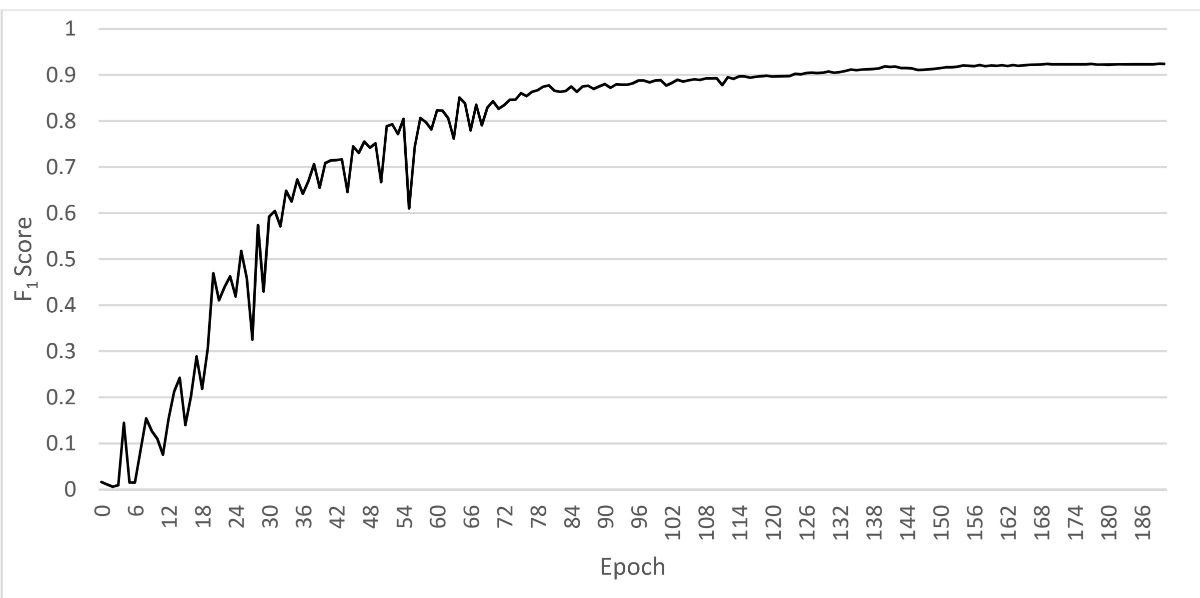

Figure 14. Performance matrices chart

TABLE II. COMPARATIVE ANALYSIS OF EXISTING METHODS

Ref.	***Year***	***Method and Algorithm***	***Accuracy***	***Limitation***
[19]	2023	Mask R-CNN, Google Translation API, gTTS API	83.9% mAP	Requires significant computational resources and lacks real-time video stream detection
[20]	2023	OCR, Text-to-Speech, Google Assistant, LSTM, Image Processing	High (specific accuracy not provided)	Needs improvements for mass production and cost-comparative analysis
[21]	2023	YOLOv3, COCO dataset, Euclidean distance calculation	High (specific accuracy not provided)	Limited by the resolution of input images, high computational requirements
[22]	2022	CNN, LSTM, Google Text-to-Speech API	83%	High computational resources required, limited real-time video processing capability
[23]	2022	YOLOv3, COCO dataset, Google Text-to-Speech API	90%	High computational resources required, dependent on internet speed for GTTS API
[24]	2022	YOLOv5, GPS-based smart stick, 3D-designed wearable model	82.33% for wearable model, 96.14% for smart stick, 89.24% overall	High computational resources required, dependent on the performance of low-cost hardware
[25]	2022	Low-cost photo cameras, LIDAR sensor, HQ camera, Machine Learning techniques	High (specific accuracy not provided)	High computational cost, limited by the performance of low-cost hardware
[26]	2022	YOLO algorithm, OpenCV library, Raspberry Pi	High (specific accuracy not provided)	Limited to the objects YOLO algorithm was trained on, requires hardware setup

9. CONCLUSION

The proposed system has demonstrated exceptional performance in both detection and navigation modes. Through rigorous testing, it has been found that the detection mode consistently provides accurate identification of objects with a high level of precision, due to the high accuracy of the system reaching 92%, which indicates high performance in every category. This ensures that the system can effectively recognize and respond to obstacles in its environment. Besides, the navigation mode algorithm implemented is found delivering obstacle avoidance instruction effectively in various arena type conditions; ensuring a smooth obstacle-free movement. In general, the system has demonstrated an impressive success in the area of both detection and navigation, in terms of supporting especially safety and efficiency for blind and vision impaired.

REFERENCES

[1] L. K. Romo, C. Alvarez, and M. J. Taussig, "An examination of visually impaired individuals' communicative negotiation of face threats," J. Soc. Pers. Relat., vol. 40, no. 1, pp. 152–173, Jan. 2023, doi: 10.1177/02654075221114048.

[2] B. Deepthi Jain, S. M. Thakur, and K. V Suresh, "Visual Assistance for Blind Using Image Processing," in 2018 International Conference on Communication and Signal Processing (ICCSP), IEEE, Apr. 2018, pp. 0499–0503. doi: 10.1109/ICCSP.2018.8524251.

[3] V. Bobby, S. Praneel, G. K. S, A. A. S, and R. S. L, "EBICS: Even Blind I Can See, A Computer Vision Based Guidance System For Visually Impaired," in 2023 International Conference on Control, Communication and Computing (ICCC), IEEE, May 2023, pp. 1–6. doi: 10.1109/ICCC57789.2023.10165192.

[4] R. Ghatkamble, K. Ratish Kumar, S. John Hrithik, J. Harshith Kumar, and P. S. Sujan, "Computer Vision and IoT-Based Smart System for Visually Impaired People," in 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), IEEE, Apr. 2023, pp. 1–5. doi: 10.1109/ICAIA57370.2023.10169589.

[5] R. A. Alqahtani, R. S. Alqahtani, Q. A. Almutairi, R. M. Aloqayli, and H. H. Alkasem, "Ayn: Assistive Tool to Describe Images for Blind and Visually Impaired," in 2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU), IEEE, Mar. 2023, pp. 25–30. doi: 10.1109/WiDS-PSU57071.2023.00018.

[6] P. Bharathi, B. L. D. Spurthy, A. Bhavani Priya, B. Yashvardhan Krishna Murthy, and CH. Charan Sai Varaha Teja, "Real- Time Object Detection for Blind People," Int. J. Adv. Res. Sci. Commun. Technol., pp. 262–266, Apr. 2023, doi: 10.48175/IJARSCT-9354.

[7] P. B, C. V. Josphine, N. Jenev J, A. G, C. L. K. R, and V. G, "Robotic Accident Prevention and Alert System for Visually Impaired," in 2023 International Conference on Disruptive Technologies (ICDT), IEEE, May 2023, pp. 672–675. doi: 10.1109/ICDT57929.2023.10150923.

[8] A. El sheikh, "Review of Robotics Systems Available to aid Visually Impaired Persons," Nafath, vol. 7, no. 19, Jan. 2022, doi: 10.54455/MCN.19.04.

[9] L. Pari and C. Mamani, "Development of a robotic application with voice to guide vision impairment people indoor environments based on Robotic Operating System," in 2021 IEEE International Conference on Automation/XXIV Congress of the Chilean Association of Automatic Control (ICA-ACCA), IEEE, Mar. 2021, pp. 1–7. doi: 10.1109/ICAACCA51523.2021.9465292.

[10] M. Ahmed, K. A. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal, "Survey and Performance Analysis of Deep Learning Based Object Detection in Challenging Environments," Sensors, vol. 21, no. 15, p. 5116, Jul. 2021, doi: 10.3390/s21155116.

[11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," pp. 7464–7475, 2023, doi: 10.1109/cvpr52729.2023.00721.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

[13] Y. Zhang, Y. Sun, Z. Wang, and Y. Jiang, "YOLOv7-RAR for Urban Vehicle Detection," Sensors, vol. 23, no. 4, p. 1801, Feb. 2023, doi: 10.3390/s23041801.

[14] G. Yan et al., "A Flame Detection Algorithm Based on Improved YOLOv7," Appl. Sci., vol. 13, no. 16, p. 9236, Aug. 2023, doi: 10.3390/app13169236.

[15] A. Thakuria and C. Erkinbaev, "Improving the network architecture of YOLOv7 to achieve real-time grading of canola based on kernel health," Smart Agric. Technol., vol. 5, no. August, p. 100300, 2023, doi: 10.1016/j.atech.2023.100300.

[16] R. K. Megalingam, V. Shriram, B. Likhith, G. Rajesh, and S. Ghanta, "Monocular distance estimation using pinhole camera approximation to avoid vehicle crash and back-over accidents," in 2016 10th International Conference on Intelligent Systems and Control (ISCO), IEEE, Jan. 2016, pp. 1–5. doi: 10.1109/ISCO.2016.7727017.

[17] A. Setiawan, H. Hadiyanto, and C. E. Widodo, "Distance Estimation Between Camera and Shrimp Underwater Using Euclidian Distance and Triangles Similarity Algorithm," Ingénierie des systèmes d Inf., vol. 27, no. 5, pp. 717–724, Oct. 2022, doi: 10.18280/isi.270504.

[18] V. S. Sadanand, K. Anand, P. Suresh, P. Kadyada Arun Kumar, and P. Mahabaleshwar, "Social distance and face mask detector system exploiting transfer learning," Int. J. Electr. Comput. Eng., vol. 12, no. 6, p. 6149, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6149-6158.

[19] N. Alzahrani and H. H. Al-Baity, "Object Recognition System for the Visually Impaired: A Deep Learning Approach using Arabic Annotation," Electronics, vol. 12, no. 3, p. 541, Jan. 2023, doi: 10.3390/electronics12030541.

[20] A. A. Llorca, "AI-WEAR: SMART TEXT

READER FOR BLIND/VISUALLY IMPAIRED STUDENTS USING RASPBERRY PI WITH AUDIO-VISUAL CALL AND GOOGLE ASSISTANCE," Int. J. Adv. Res. Comput. Sci., vol. 14, no. 03, pp. 119–129, Jun. 2023, doi: 10.26483/ijarcs.v14i3.6997.

[21] V. E. CHANDU, C. HARSHITHA, V. N. SREE, and E. GANESH, "OBJECT DETECTION AND ESTIMATING THE DISTANCE BETWEEN DETECTED OBJECTS USING DEEP LEARNING ALGORITHMS," INTERANTIONAL J. Sci. Res. Eng. Manag., vol. 07, no. 04, Apr. 2023, doi: 10.55041/IJSREM18689.

[22] J. Ganesan, A. T. Azar, S. Alsenan, N. A. Kamal, B. Qureshi, and A. E. Hassanien, "Deep Learning Reader for Visually Impaired," Electronics, vol. 11, no. 20, p. 3335, Oct. 2022, doi: 10.3390/electronics11203335.

[23] M. I. T. Hussan, D. Saidulu, P. T. Anitha, A. Manikandan, and P. Naresh, "Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People," Int. J. Electr. Electron. Res., vol. 10, no. 2, pp. 80–86, Jun. 2022, doi: 10.37391/ijeer.100205.

[24] N. Kumar and A. Jain, "A Deep Learning Based Model to Assist Blind People in Their Navigation," J. Inf. Technol. Educ. Innov. Pract., vol. 21, pp. 095–114, 2022, doi: 10.28945/5006.

[25] C. Dragne, I. Todiriţe, M. Iliescu, and M. Pandelea, "Distance Assessment by Object Detection—For Visually Impaired Assistive Mechatronic System," Appl. Sci., vol. 12, no. 13, p. 6342, Jun. 2022, doi: 10.3390/app12136342.

[26] M. Kadhim and B. Oleiwi, "Blind Assistive System based on Real Time Object Recognition using Machine learning," Eng. Technol. J., vol. 40, no. 1, pp. 159–165, Jan. 2022, doi: 10.30684/etj.v40i1.1933.

Mohanad A. Kadhim
Received B.Sc. In Computer Science degree from Babylon University, Babylon, Iraq.

Nidhal K. El Abbadi
Prof. Nidhal El Abbadi is currently a professor of computer techniques engineering, Al-Mustaqbal university. He graduated with a Ph.D. degree in computer science from Informatics Institute for Postgraduate Studies (IIPS) – Iraq Commission for Computer and Informatics 2006. Earlier, in 2002 he completed his M.Sc. degree in Computer Science from (IIPS). He received his B.Sc. (Hons) degree in Computer Science from Al-Mustansiriya University in 2000, and received his B.Sc. in Chemical Engineering, from the University of Baghdad in 1976. He worked as an engineer in many industrial plants. In 2006, Nidhal joined the academic staff at Al-Mustansiriya University. His research area includes image processing, computer vision, steganography, and IoT (ACM Classification Codes (I3, I4, I5)). He has successfully supervised many Ph.D. and master's students and published more than 300 articles in various international journals and conference proceedings. He published eight books related to computer science. He acts as a reviewer for various journals and conferences. He has also served as a conference chair, and as a technical committee for numerous international conferences.