A System for the Detection and Identification of Objects and Their Distances to Aid Blind People

Abdelfatah A. Tamimi, Alaa A. Abu Hammad, Ayman M. Abdalla, Omaima N. Al-Allaf
Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan, Amman, Jordan

Abstract - *This paper presents a novel system to help blind people detect and identify objects and their locations. This system represents visual objects with audio through a simulator to inform blind people about objects' names and locations. The system captures images with a portable camera device. Then, the system obtains the location of each object through a fixed indoor position system. In the next step, the system employs real-time image recognition to detect and identify objects. After that, object distances from the person are computed. Existing objects inside the room are stored in a Reference Database using an Indoor Positioning System. For object detection and identification, the input images are compared with the objects in the Reference Database. Finally, the names and locations of the recognized objects are read audibly through speakers. Test results showed the system's effectiveness in recognizing and identifying commonly used objects and computing their distances.*

Keywords: Image detection, Image identification, IPS, SIFT, SURF

1 Introduction

Recent studies have shown that at least 217 million people have moderate to severe vision impairment and 36 million people are blind [1]. In addition, more than 826 million people live with near-vision impairments [2]. These visually impaired people, especially the blind, generally face burdens in finding their way through unknown places and have to overcome further difficulties to recognize objects and people and their locations.

This paper provides an analysis and evaluation for the technologies used in the object identification task, and then proposes a system for the Detection and Identification of Objects and their Distances to aid Blind people (DIODB) through image to sound conversion. The proposed DIODB system uses the Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) robust algorithms in the detection and recognition of objects in images. These methods provide more matching, good processing speed, and robustness to variations with respect to illumination, rotation, and scaling to identify objects. Then, the output of these methods are processed further for converting the information of the recognized object to the speech form. Feature-detection, matching, and mosaicking with SIFT and SURF was discussed and implemented in details by [3].

DIODB employs an Indoor Positioning System (IPS) to locate objects inside the room. IPS works in a way similar to Global Positioning System (GPS), but rather than utilizing a satellite signal, IPS exchanges signals between the location devices and sensors of smart devices.

2 Related work

Since the 1970s, systems based on portable cameras have been evolving considerably, continuously creating new tools for the visually impaired [4]. The visual world is very complex, as objects may be added or relocated within the visual field at any time. It is difficult to represent objects mentally in the visual world because of world and object complexity. Each object comprises many features such as line edges, orientations, colors, luminance intensities, and moving parts. Although many applications were developed for this task, such applications are often not very robust because machine vision does not offer a robust and actual representation of the real world [5]. For example, biometric recognition modalities may be used in identifying people [6,7,8,9,10] where different methods are needed for object identification [4].

A system was developed by [11] to provide safety assistance to car drivers. This system was able to recognize objects, signs, and road surface to make decisions on warning or acting on behalf of a driver. However, this system was not suited to aid walking visually impaired people, and it did not concentrate on identifying the detected objects or people.

Several general methods for object recognition were considered by [12] as they developed a novel segmentation system. An excellent real-time object detection system is named YOLO [13]. This system performs fast object-recognition with high accuracy. Even though it was not designed for blind people, it could be modified or combined with another system to suit blind people and to provide extra features such as computing distances and giving voice notifications for designated objects.

A method for object and scene detection for blind peoples using vocal vision was presented by [14], but implementation results were not discussed and the system's effectiveness was not demonstrated. Reference [15] developed an automatic algorithm for object recognition and detection based on Affine SIFT (ASIFT) key points. They presented an ASIFT method to identify objects with full boundary detection by combining a fixed constant scale and a zone integration algorithm. A strong integration algorithm is used in the region to identify the object with full boundaries and detect it in other images based on the Aseptic key points and measure the similarity of the areas merged into the image. Their experimental results showed that this method was effective in recognizing and detecting the object.

Reference [16] proposed a system to provide blind people with real-time visual recognition to transform the visual world into 3D audio to inform blind people of names of objects and their locations.

Modern mobile phones have strong communication capabilities and are equipped with different types of integrated sensors for various functions. A survey of systems for finding indoor position of a person using Wi-Fi and a smartphone is available [17]. These systems are not sufficient for aiding visually impaired people find their way and they do not detect objects and people unplugged from the system. However, such systems may be used in conjunction with recognition and identification methods to provide the visually impaired with a more helpful system.

Reference [18] suggested that IPS uses multiple signals from location devices in addition to sensors including motion sensors to accurately calculate user position. Some IPS systems analyze the system's underlying data generated while the system is running. These processes can predict future user traffic even in multi-story buildings [18].

Reference [19] proposed an object recognition method to help blind people find missing items using SURF. The proposed recognition process matches the individual features of the user-queried object to a database of features that contain different pre-saved personal items. The matching was performed under various conditions including image scale, partial occlusion, translation, rotation, and change in viewpoint.

3 Methodology of proposed system

3.1 System overview

The proposed DIODB system identifies objects and computes their distances from the user, and then it relays this information by sound to benefit visually impaired users. The system starts with activating the portable camera and capturing an image of the scene in which objects are to be discovered. Then, the system analyzes this image by specifying feature points that identify key properties of this image. These feature points are used in image matching with reference images previously stored in the DIODB database. This is performed by finding the common characteristics between the scanned image and reference images. If a matching is found for three or more feature points of the scanned image, the system will detect the object. After the process of matching and detecting objects in the scene image, the system copies the matched images from the database to a specific folder of discovered images and then stores the names of the recognized objects in a text file. Finally, the system will read this text file audibly. Next, the IPS system responsible for locating objects in the closed place starts to locate the positions of the discovered objects and stores their coordinates. After that, the system will compute the distances to these objects and outputs them audibly through the speakers. Figure 1 shows a general diagram of the system.

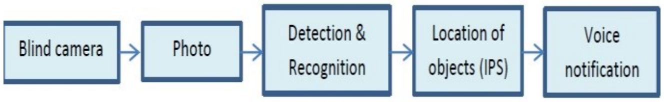

Figure 1: A general diagram of the DIODB system.

3.2 Phase 1: Scene shooting

Figure 2 illustrates the scene shooting phase of the DIODB system. This phase starts with shooting a given number (J) of images of the scene with a portable camera and saves them into the testing database. Figure 3 shows an example of a scene obtained at the scene shooting phase.

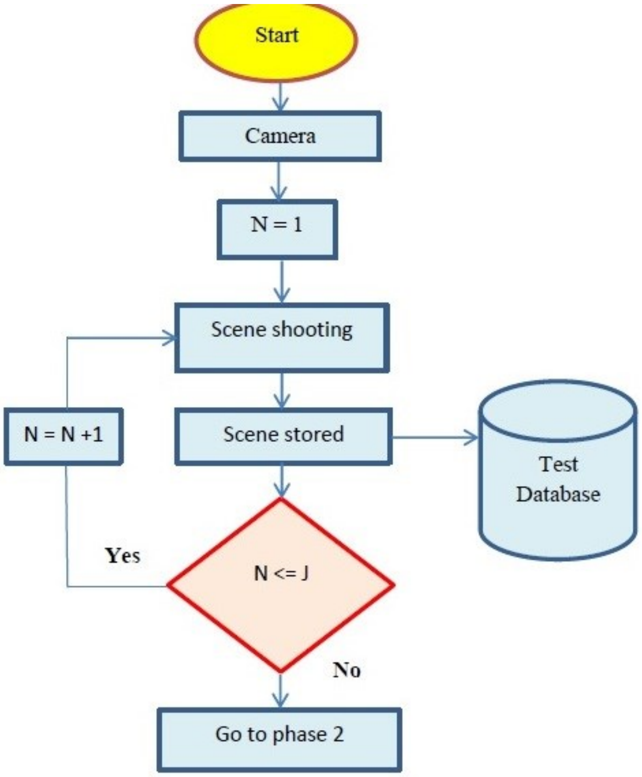

Figure 2: Flowchart of the scene shooting phase.

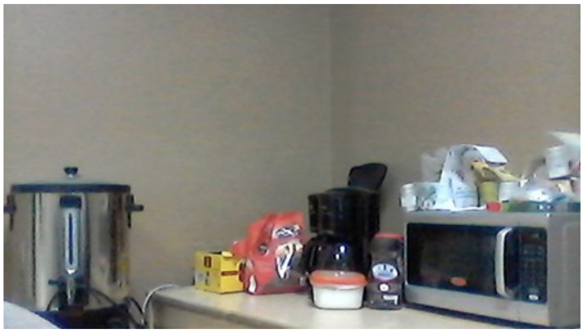

Figure 3: An example of a scene obtained at the scene shooting phase.

3.3 Phase 2: Object detection and recognition

A flowchart of the object detection and recognition phase is shown in Figure 4. After obtaining J images of the scene in Phase 1, Phase 2 detects and recognizes the objects in the scene to match them with the objects in the Reference Database. When implemented, DIODB started with extracting 300 key points for each image stored in the database of reference images. Figure 5 shows an example of a scene image with 300 feature points. Then, DIODB selected the strongest 100 key points for each of these objects. Figure 6 shows the box image from the Reference Database after selecting the 100 strongest feature points.

The algorithm tries to match the objects in the scene image with those in the reference images database by making a linear connection between the key points of each object in the scene and the reference data. When three or more links are found between a scene object and a reference object, the object is identified. Figure 7 shows matched points of the object identified and shown in Figure 8.

After objects' detection, the IPS system is applied to determine the location of the objects in the room. The final step is to convert all objects' names and locations to voice by an existing text-to-speech conversion system.

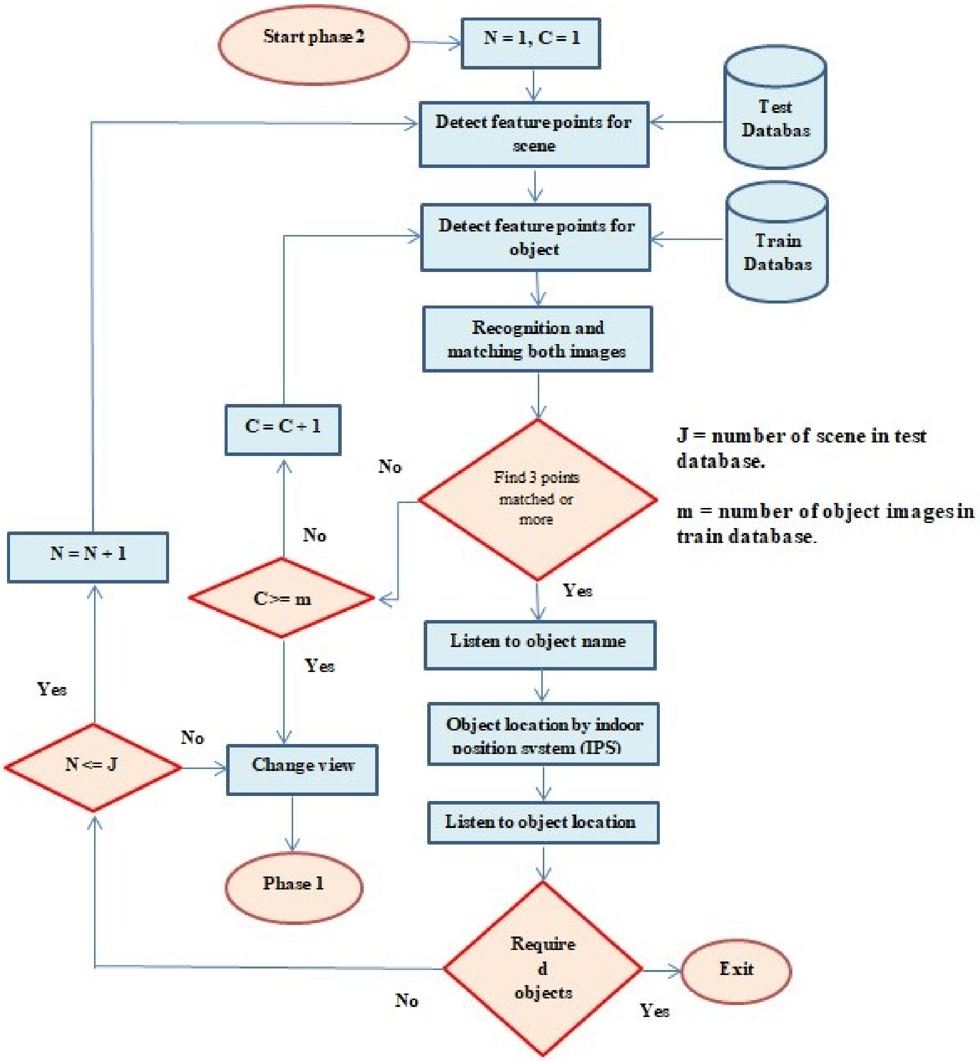
Start phase 2
N = 1, C = 1
Detect feature points for scene
Test Databas
Detect feature points for object
Train Databas
Recognition and matching both images
J = number of scene in test database.
m = number of object images in train database.
C = C + 1
No
Find 3 points matched or more
No
N = N + 1
C>= m
Yes
Listen to object name
Yes
Yes
Object location by indoor position system (IPS)
No
N <= J
Change view
Listen to object location
Phase 1
Require d objects
Exit
No
Yes

Figure 4: Object detection and recognition phase.

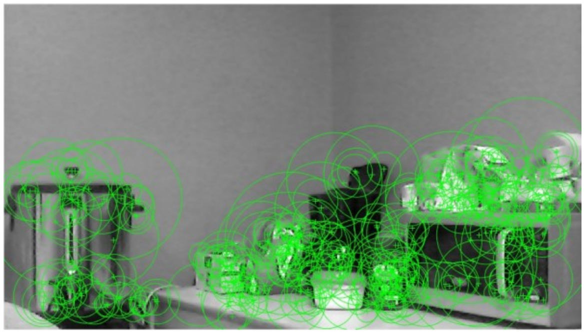

Figure 5: An example of a scene image with 300 feature points.

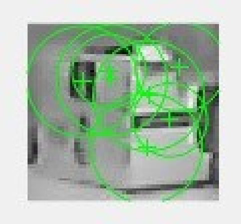

Figure 6: An example image of a reference object with 100 strongest feature points.

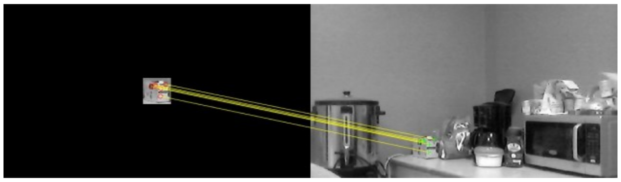

Figure 7: Matched points.

Figure 8: Identified object.

3.3.1 Indoor position system

In DIODB, we replace the sensors in an indoor position system (IPS) with cameras having IPS capabilities located inside the room. When taking a picture of objects through the camera on the blind glasses, the cameras inside the room take a full picture of the room. At the same time, processing the image begins to identify the objects in the room and locate them using the image taken by the cameras inside the room and the IPS. Once the objects have been detected, identified and positioned, their locations are convert to voice. The blind person can be directed to the intended object using the building map and the IPS.

Steps for determining object distances using IPS in the DIODB system:

1. Initialize each object's location (x_1,y_1,z_1) by using fixed IPS cameras, where z_1 = object height from the ground.
2. The user enters the room with an IPS-enabled camera to locate his position.
3. The user searches for needed items by executing Phase 1 and Phase 2.
4. If object (x_1,y_1,z_1) is detected and recognized, the IPS system will compute its distance from user's location (x_2,y_2,z_2) with (1).

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (1)$$

Figure 9 shows initialized coordinates for all objects' locations on the map of the room by using IPS.

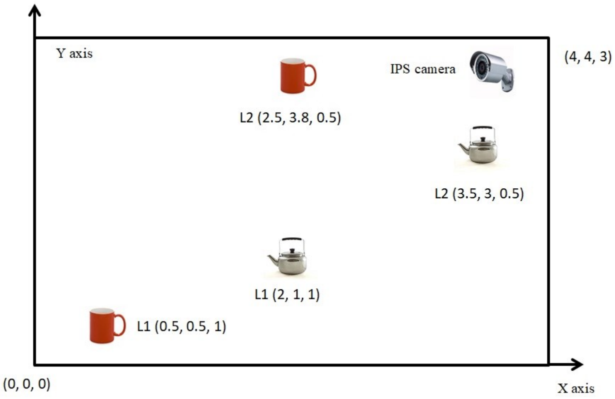

Figure 9: Initialize coordinates for all objects location .

3.3.2 Voice notification

In this phase, the system gives information about the name and location of the object by playing the audio files for the blind person and directing that person through IPS to reach the required object.

4 Implementation results and analysis

The program was executed and samples were taken from different distances and angles. The number of objects detected, the number of matched objects, and the accuracy calculations were computed for matching the objects. The following assumptions were made:

1. The features of the camera do not have accuracy zooming.
2. There is no representation from angle 90° on the Y axis vertically.

3. The amount of lighting should be taken into account.

To choose an appropriate value of the number of captured input scene images (J), the DIODB algorithm was tested with values of J from 1 to 5. As shown in Table 1, the values ($J \geq 3$) gave the maximum number of recognized images, so ($J = 3$) was the optimal number.

Table 1. Multiple trials for J = 1 to 5.

No. of Scenes	No. of Objects	No. Recognized
1	6	1
2	6	3
3	6	5
4	6	5
5	6	5

Table 2 shows a general comparison between DIODB and some related systems in terms of the features they provide, where DIODB appears to offer relatively good features for users. More features could be added to the system in future work. The test results of [19], shown in Table 3, showed that the overall percentage of detected elements was 84%.

In the DIODB system, the first step is to detect objects in images by using the speed up robust feature (SURF) algorithm for one or more objects that have multiple images in different scenes. Object detection in DIODB uses SURF as a robust speed up feature. Every object has 10 2D images where these images were taken from multi-view scenes.

Table 2. General comparison between DIODB and some related systems.

Systems	Camera	Detection	Recognition	IPS	Ultrasonic sensor	Voice	Vibration	Distance
Electronic Travel Aids	No	No	No	No	Y	No	Y	No
Augmented Reality	Y	No	No	No	No	No	No	Y
Navigation Assistance for Visually Impaired	Y	Y	Y	No	Y	Y	No	Y
Deep-See	Y	Y	Y	No	No	Y	No	No
DIODB	Y	Y	Y	Y	No	Y	No	Y

Table 3. Recognition accuracy for reference objects by [19].

Objects	No. of Images	Correctly Detected	False Positive	Accuracy
Keys	20	19	0	95%
Cell	20	12	0	60%
Wallet	20	17	0	85%
Sunglasses	20	17	0	85%
Sneaker	20	19	0	95%
Total	100	84	0	84%

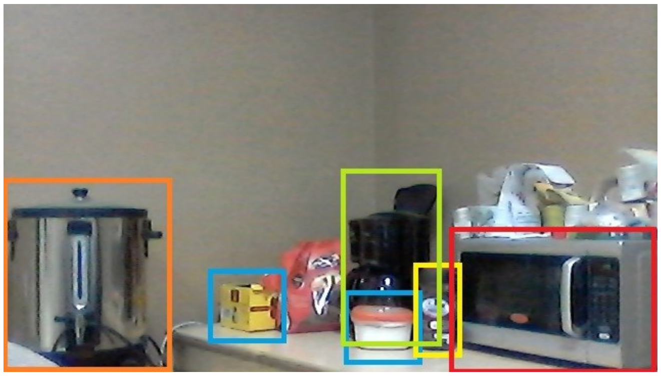

Figure 10: Object detection by DIODB.

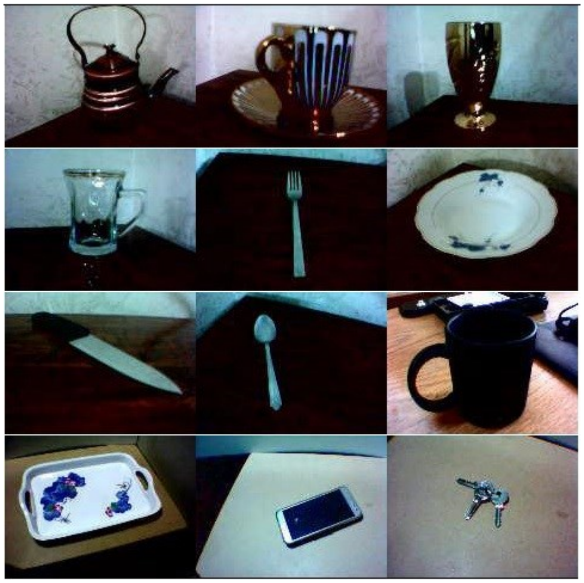

Figure 11: Sample Reference Database Images.

Figure 10 shows object detection using one image of one scene as input to our system. Using the SURF algorithm on this image detected each object by itself and stored the data in a table.

Figure 12: A Selected Test Image.

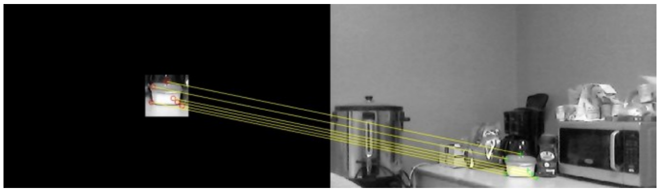

Figure 13: Identifying an Object.

Figure 14: Identified Object.

Object recognition is the process of identifying an object in a digital image or video. The proposed DIODB system creates a reference database that has 510 images for many objects with different poses for each object. The Reference Database is used in the learning phase of the object recognition. In addition, a Test Database was used for comparing the results with images in the Reference Database. The Test Database images were not identical to the images in the reference database, but generally similar object were recognized and unknown objects were not. Figure 11 shows a sample of selected images from the Reference Database where Figure 12 shows a selected test image.

To recognize the test image, the system uses the geometric transformations method in SURF and compares this image with images in the Reference Database to be identified. As shown in Figure 13, a linear mapping is made between the identified object in the scene (on the right-hand side of Figure 13) and the object (on the left-hand side of Figure 13) coming from the Reference Database. The identified object, which was extracted from the scene in Figure 13, is shown in Figure 14.

Tables 4, 5, and 6 are multiple results with different objects and with angles (ax) degrees on the x-axis, (ay) degrees degree on the y-axis, and (d) distances in centimeters. As seen in these tables, the DIODB system was able to recognize objects with relatively high accuracy, even with rotated objects.

5 Conclusions

In this research, a new system employing the indoor position system was introduced to identify objects and their locations and to convert obtained information into audio. Implementation results showed that this system could help the visually impaired to track various indoor objects. The system could be enhanced by identifying objects from multi-view scenes and accounting for more factors such as occlusion and distortion. More features could be added to enhance the system's usability. Extended future work may use video to account for object's motion characteristics where it could benefit from existing software such as YOLO [13]. Related future work may combine this system with augmented reality software such as Google Tango and ARCore.

Table 4. (d = 25 cm, ax = 0, ay = 0).

No. objects	Detection	Recognition	Accuracy
1	1	1	100%
2	2	2	100%
5	5	5	100%
10	10	9	90%
15	15	13	86%
20	20	19	95%

Table 5. (d = 50 cm, ax = 0, ay = 0).

No. objects	Detection	Recognition	Accuracy
1	1	1	100%
2	2	2	100%
5	5	5	100%
10	10	8	80%
15	15	12	80%
20	20	17	85%

Table 6. (d = 100 cm, ax = -90, ay = 30).

No. objects	Detection	Recognition	Accuracy
1	1	1	100%
2	2	2	100%
5	5	4	80%
10	10	8	80%
15	15	10	67%
20	20	14	70%

6 References

[1] R.R.A. Bourne, S.R. Flaxman, T. Braithwaite, M.V. Cicinelli, A. Das, J.B. Jonas, et al. "Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis"; Lancet Glob Health, Vol. 5, No. 9, 888–897, Sep 2017.

[2] T.R. Fricke, N. Tahhan, S. Resnikoff, E. Papas, A. Burnett, M.H. Suit, T. Naduvilath, K. Naidoo. "Global Prevalence of Presbyopia and Vision Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling"; Ophthalmology, May 2018.

[3] Shaharyar A.K. Tareen and Zahra Saleem. "A Comparative Analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK"; International Conference on Computing, Mathematics and Engineering Technologies – iCoMET, 2018.

[4] Kirti P. Bhure and J.D. Dhande. "Object Detection Methodologies for Blind People"; International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 5, Issue 1, 194-198, 2017.

[5] Mitchell. R.P. Lapointe and Bruce Milliken. "Influences of Context on Object Detection and Identification in Natural Scenes"; Ph.D. Thesis, McMaster University, 2016. http://hdl.handle.net/11375/20265

[6] O.N. Al-Allaf, A.A. Tamimi and M.A. Alia. "Face Recognition System Based on Different Artificial Neural Networks Models and Training Algorithms"; International Journal of Advanced Computer Science and Applications, Vol. 4, No. 6, 40-47, 2013.

[7] M.A. Alia, A.A. Tamimi and O.N. Al-Allaf. "Integrated System for Monitoring and Recognizing Students During Class Session"; International Journal of Multimedia & Its Applications, Vol. 5, No. 6, 45-52, 2013.

[8] D.R. Ibrahim, A.A. Tamimi and A.M. Abdalla. "Performance Analysis of Biometric Recognition Modalities"; International Conference on Information Technology (ICIT'2017), pp. 980-984, Amman, Jordan, 17-18 May 2017.

[9] A.A. Tamimi, O.N. Al-Allaf and M.A. Alia. "Real-Time Group Face-Detection for an Intelligent Class-Attendance System"; International Journal of Information Technology and Computer Science, Vol. 6, pp. 66-73, 2015. doi: 10.5815/ijitcs.2015.06.09.

[10] A. Tamimi, O.N.A. Al-Allaf and M.A. Alia. "Eigen Faces and Principle Component Analysis for Face Recognition Systems: A Comparative Study"; International Journal of Computers & Technology, Vol 14, No. 4, 5650-5660, Feb. 2015.

[11] Meiyuan Zhao. "Advanced Driver Assistance Systems"; Technical Paper, Security & Privacy Research, Intel Labs, 2015. https://www.semagarage.com/assets/pdf/advanced-driver-assistant-system-paper.pdf.

[12] Kiana Ehsani, Roozbeh Mottaghi and Ali Farhadi. "SeGAN: Segmenting and Generating the Invisible"; 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-22 June 2018. https://dx.doi.org/10.1109/CVPR.2018.00643.

[13] https://pjreddie.com/darknet/yolo/

[14] Nalawade M. Rajendra, Wagh Vrushali and Kamble Shradha. "An Approach for Object and Scene Detection for Blind Peoples Using Vocal Vision"; Int. Journal of Engineering Research and Applications, Vol. 4, Issue 12, Part 5, 1-3, Dec. 2014.

[15] Reza Oji, "An Automatic Algorithm for Object Recognition and Detection Based on ASIFT Key Points"; Signal & Image Processing: An International Journal, Vol. 3, No. 5, 29-39, 2012.

[16] Rui Jiang, Qian Lin and Shuhui Qu. "Let Blind People See: Real Time Visual Recognition with Results Converted to 3D Audio"; Technical Report, Stanford University, 2016. http://cs231n.stanford.edu/reports/2016/pdfs/218_Report.pdf

[17] Gajanan D. Bonde, Pooja U. Barwal, Sandhya R. Pal, Sumaiya I. Khan and Kiran Ablankar. "Finding Indoor Position of Person Using Wi-Fi & Smartphone: A Survey"; International Journal for Innovative Research in Science & Technology, Vol. 1, Issue 8, January 2015.

[18] Marin Kaluža, Kristina Beg and Bernard Vukelic. "Analysis of an Indoor Positioning System"; Zbornik Veleučilišta u Rijeci, Vol. 5, No. 1, 13-32, 2017.

[19] Ricardo Chincha and YingLi Tian. "Finding Objects for Blind People Based on SURF Feature"; 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops, Atlanta, GA, USA, 12-15 Nov. 2011.