International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15 (2023) pp. 641-650

Received: 10 February 2023; Accepted: 21 May, 2023; Published: 8 October, 2023

Novel Approach for Easy Navigation based on Acoustics and Relative Senses for Visually Impaired People

Prajakta S. Saraf[1], Sanika A. Watve[2] and Anagha R. Kulkarni [3]

[1] MKSSS's Cummins College of Engineering for Women Pune
prajakta.saraf@cumminscollege.in

[2] MKSSS's Cummins College of Engineering for Women Pune
sanika.watve@cumminscollege.in

[3] MKSSS's Cummins College of Engineering for Women Pune
anagha.kulkarni@cumminscollege.in

***Abstract*: People with blindness or partial-sightedness have a decreased or absolutely no ability to visualize the outside world. These people experience limitations in their mobility, productivity, and independence which increases the risk of injuries and accidents. Visually impaired people can navigate in their houses without assistance since they are familiar with their surroundings. However, navigating around new locations is the most difficult task for visually impaired people. This paper focuses on designing a novel system for the easy navigation of visually impaired people in familiar and unfamiliar surroundings. This paper enhances the features of the system by using novel techniques that in turn employ a blend of Deep Learning algorithms with simple coordinate geometry. The aim of this paper is to provide a generic, multipurpose system for the visually impaired or partially sighted that would help them locate objects in their surroundings. The object detection algorithm used in this piece of work, renders an accuracy of 99.5%.**

***Keywords*:** Coordinate Geometry; Deep Learning; Navigation based on Acoustics and Relative Senses; Object Detection; Visually impaired (VI).

I. Introduction

A person's primary sensory organ is their eyes. Train station itineraries, signs indicating the right route or a possible threat, and a billboard marketing a new service are all instances of visual information that we see on a frequent basis. Because information access is synonymous with autonomy, most of this information is inaccessible to the VI people, restricting their freedom, independence, and autonomy [1, 2, 3].

As stated by WHO, almost more than 2.2 billion individuals worldwide have been affected by vision impairment. Low and middle-income countries are projected to have four times the occurrence of distant sight problems as high-income countries when it comes to geographic disparities. Visual impairment is estimated to influence well over 80% of the population in western, eastern, and central sub-Saharan Africa, but less than 10% of men in high nations like North America, Australasia, Western Europe, and Asia-Pacific.

For someone who is differently abled, true independence and accessibility is crucial [4]. A blind person can live independently with the help of specially made adapted goods. There are numerous adaptive tools that can assist a VI person in leading a normal life, but they are difficult to find in local stores or markets.

Blindness and impaired vision necessitate the development of automated processes to aid those who are VI [5]. The goal of this paper is to develop a system that will support VI persons in traversing roads.

This paper intends to describe a system in detail that would in return help the VI to navigate in outdoor as well as indoor environments without assistance. This adaptability to the surrounding environment and independence to live and move about can be achieved with the aid of the Captioning system is discussed in this paper.

Next section presents literature review. Methodology and results are discussed in sections 3 and 4. Section 5 presents a conclusion. Future work is explained in section 6.

II. Literature Review

A. Existing systems both hardware and software.

As per prior arts, there have been many systems developed for object detection to recognize the objects in an image for VI. Few of those systems are as follows:

Initiated on October 11, 2012, TapTapSee[34] is a mobile camera application powered by the CloudSight Image Recognition API that is specifically made for blind and VI people. TapTapSee uses the camera and VoiceOver capabilities of your device to take pictures or videos of anything and read its name aloud to you. Take pictures by double-tapping the right or left sides of the screen, respectively, or record videos by doing the same. Any two- or three-dimensional object can be precisely analysed and identified by TapTapSee[34] in a matter of seconds from any perspective. The identifier is then read aloud by the device's VoiceOver.

The free smartphone software Lookout by Google, which was made available for Android smartphones in March 2019,has the ability to automatically read and scan text, identify products, and describe things. Photographs mode (beta) can provide a summary and additional information about still images. Improved reading order for menus, receipts, and other formatted text is available in Text and Documents mode. Lookout identifies items more precisely in Explore mode. Lookout detects many more food items in Brazil and India when it is in Food Labels mode. This system's lack of voice commands and the inability to tap anywhere on the screen to access the system's features are also drawbacks

Mediate, a Boston-based MIT spinoff AI start-up, is creating SuperSense. The most intelligent assistive programme, Supersense, enables blind and VI users to read, discover items, and autonomously explore new environments. To improve accessibility for the blind and low vision communities, it offers a set of digital eyeballs. Once more, this system's drawbacks include the lack of voice instructions andthe fact that it is not tap-based.

Be My Eyes [35] is a free app that establishes live video calls between blind and low-vision users and sightedvolunteers and company representatives to provide visualassistance. In order to help blind and low-vision people live more independent lives, sighted volunteers donate their eyes every day to accomplish chores big and little.

Due to its portability and low cost, the white cane has historically been the most widely used and basic tool for obstacle detection. It gives users the ability to accurately scan the environment in front of them and find ground impediments like holes, steps, walls, uneven surfaces, downstairs, etc.
When a user presses keys on a mobile device, a technology called "Roshini" uses audio messages to guide them through the building. It mounts ultrasonic modules on the ceiling at regular intervals to use sonar technology to determine the user's location. This system is portable, simple to use, and unaffected by changes in the environment.

The user's location within the structure is determined usingthe "Roshini" system, which also allows for mobile unit navigation via audio messages. Through the installation of ultrasonic modules at regular intervals on the ceiling, it uses sonar technology to determine the user's location. Environmental changes have no effect on this technology, which is portable and simple to use.

The idea of a wearable jacket is also put forth. To informa user of the direction from which an impediment is coming, sonar sensors and vibrators are fastened to a jacket. For real- time navigation and obstacle avoidance, another jacket-type method is presented that combines an RGB-D camera with haptic devices. The traverse ability maps are provided toshow open and occupied (obstacle) spaces. The RGB-D camera creates depth data associated with RGB pictures. A consumer receives instructions such as "Go straight" and"Turn right" from four tiny vibration motors on a jacket.

B. Existing systems both hardware and software.

The following table lists down the algorithms used by one or more systems for the purpose of object detection along with their advantages and disadvantages.

Algorithm	**Advantages**	**Disadvantages**
RCNN[8,25,26, 27]	R-CNN makes use of a selectivesearch approach instead of sliding window technique which is an exhaustive search. So, the number of regions generated are drastically reduced to approximately 2000 per image by taking advantage of segmentation ofobjects. R-CNN makes use of transfer learning in which a pre-trained model such as AlexNet or VGGNet trained on ImageNet dataset and usesand adopts their weights. Based on the detection task, only the last fully connectedlayer is	The convolutional neural network extracted features from each proposal separately. As a result, repeated computations are possible. Despite the fact that the number of regions is reduced to 2000 when compared to the sliding window approach, R-CNN is computationally expensive.

	reinitialized. Additionally, R-CNN eliminates many straightforward negatives prior to training, which speeds up learning and reduces false positives.	
Fast RCNN[30]	Because it only conducts one convolution operation per image in order to give a feature map, this approach is faster than R-CNN. Fast RCNN uses softmax layer instead of Support vector machine which proved to be faster and is more accurate than SVM	The disadvantage of this is that region generation using selective search algorithms takes time and thus becomes the bottleneck
Faster RCNN[31]	In contrast to fast RCNN in which these generated feature maps were given to the selective search algorithm which is slow, feature maps are given to a separate CNN network for generating regions which is called region proposal network. The input image's feature map was quickly computed, and area features were extracted from the feature map, enabling the computation of feature extraction to be distributed across numerous regions.	To extract all the objects from the image using this approach, it has to run through numerous passes.
YOLO[6,7,32,33]	YOLO runs at a faster speed thus saving a lot of time. It is suitable for a real-time environment.	At the most two objects at a given location could be detected by YOLO which further complicates the detection of small objects. It also affects the detection of objects from a crowd of objects [4]. The most recent feature map is used for prediction, making it challenging to anticipate objects of diverse scales and aspect ratios [12]. It can predict only one class for the objects lying within a grid cell. Therefore, objects belonging to different classes falling in the same grid cell will be predicted as objects belonging to only one common class
SSD[30]	Suitable for real-time environment	It has a weak detection capacity of

	[4]. Has comparable accuracy to that of Faster R-CNN[4]. It is a faster method than methods based on two-shot RPN[10]. Works well withscenarios encompassing multiple scales and aspect ratios[4]. It can handle objects with various sizes [11].	small-sized objects

Table 1.Literature Review of Object Detection Algorithms

C. Limitations of state-of-the-art literature

TapTapSee: The disadvantage of this system is there are no voice commands given. The person needs to know in advance where to tap in order to access the system.

LookOut by Google: The disadvantage of this system is that there are no voice commands and there is no facility to tap anywhere on the screen in order to access the functionalities of the system (not a user-friendly UI).

SuperSense: The disadvantage of this system is there are no voice commands and also it is not tap based.

Be My Eyes: The disadvantage of this system is it makes VI people dependent on some other sighted person for assistance.

System Roshni: This system is limited only for indoor navigation because it requires a detailed interior map of the building.

White cane with sensors: By attaching sensors to the white cane, it becomes easy for the VI to detect obstacles but at the same time the cost also increases.

Thus, both the hardware and the software-based solutions exist but have some or the other limitations like damage to the hardware part, costly equipment, dependency on some other person, which if worked upon and improved can help the VI. None of the existing systems provide the directions based upon the object locations.

III. Methodology

The proposed system, a mobile app, helps the VI for better navigation in outdoor as well as indoor environments efficiently. It makes use of Artificial Intelligence, Machine Learning, and Deep Learning models. The system takes in video as well as image data as input and gives an audio response as output.

The proposed system gives accurate directions based on the object locations.
The system consists of 2 parts as described below:

A. Giving directions based on image

1) Input

The input to the system is a photo clicked by mobilecamera.

2) Output

Audio information about objects in the photo.

3) Processing

The first step is to detect the objects in the photo (or image).To accomplish this task object detection algorithms (YOLO v5) is used. The objects that could be detected are the sofa, chair, car and so on.

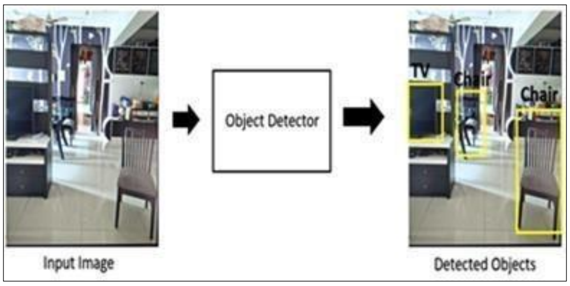


Figure 1. Objects commonly seen in a house

Figure 1 shows a few objects such as chairs, TV etc. that are commonly found in a house. The object detection algorithm detects the objects and puts a bounding box around theobject along with the class label of the detected object. The bounding box is either a rectangle or a square and the class label is the category associated with the object.

The bounding box has four coordinates as shown in figure 2:
x - upper left corner point x coordinate
y - upper left corner point y coordinate
w - width of the bounding box
h - height of the bounding box

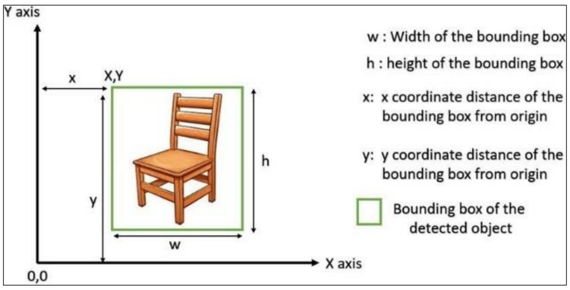


Figure 2. Coordinates of the bounding box

The centre point of every bounding box is computed (xc,yc) according to equation 1.

$$Xc = x + w \div 2 \quad (1)$$

$$Xc = x + w \div 2 \quad (2)$$

The centre point of the entire image (or photo) is computed (x_imageframe, y_imageframe) using the formula given above.

The relative position (left, right or front) of every objectwith respect to the centre of the image is calculated based on the logic depicted in figure 3.

Total number of objects lying on the left, right and front is calculated to form a summary of all the objects in the photo. This summary is converted to audio

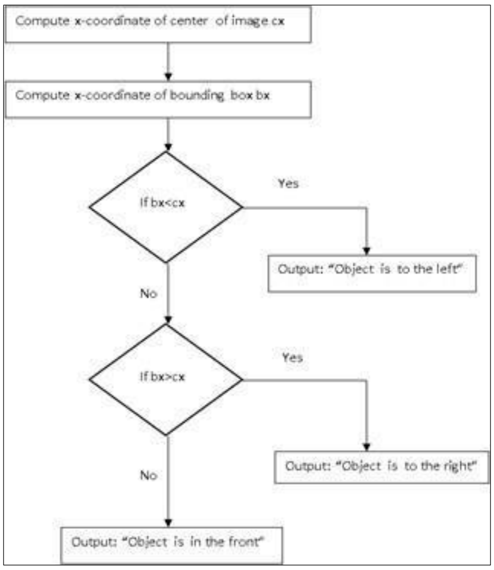


Figure 3. Flowchart for giving directions based on image data

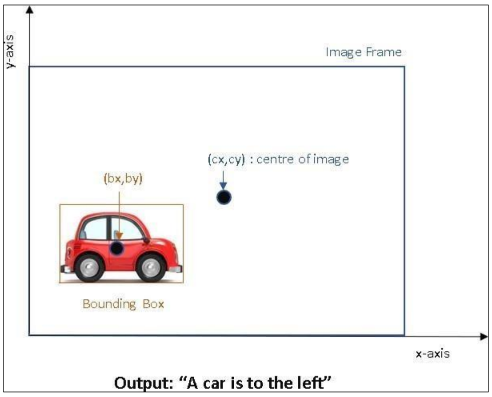


Figure 4. An image frame under computation for giving directions module

B. Detecting moving object based on Video

1) Input

The input to the system is a real time video.

2) Output

The output of the system is whether any object is approaching towards the person in audio format.

3) Processing

Video is a collection of frames. Framesrepresent the still images. The first step is to detect the objects in each frame. For this object detection algorithms YOLO are used. The detected objects are represented by the bounding box sameas in the case of the images. Algorithm discussed above is used to detect objects in every frame.

Next step is to define a threshold line. The object is said to be approaching the observer when the lower boundary line of the bounding box crosses the threshold line as shown in figure 4. When the y coordinate of the bottom line of the bounding box is less than or equalto the y coordinate of the threshold line, an approach alert isgiven. This threshold line defined is dynamic and can be changed as per the requirements of the user.

The logic used for the above mentioned purpose, is depicted in figure 5.

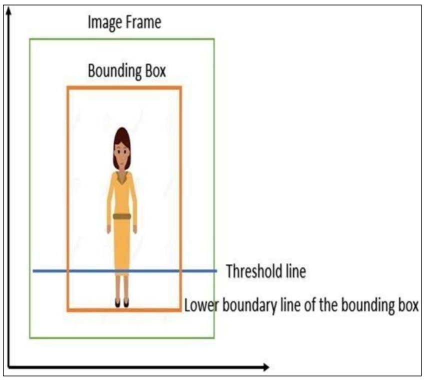


Figure 5. Threshold and lower boundary line

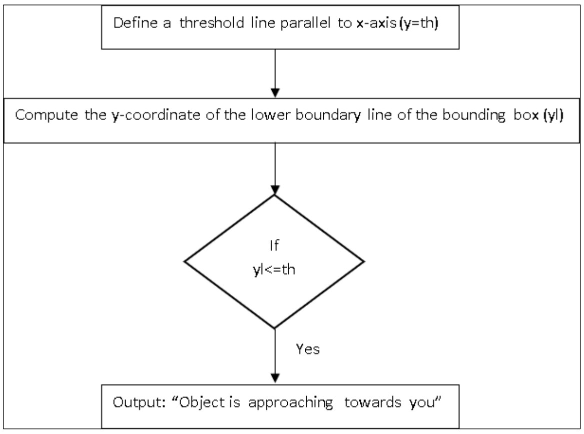
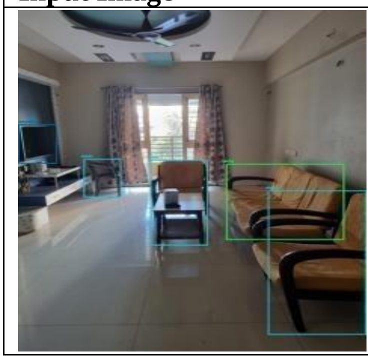


Figure 6. Flowchart of threshold boundary line

IV. Results And Analysis

A. Results

The system has been tested by deploying it as a mobile application in real time in indoor and outdoor scenarios. The following table shows the input images and the generated caption respectively.

Input Image	**Output(Audio)**
	one sofa to the right two chairs to the left one chairto the right one tv monitorto the left

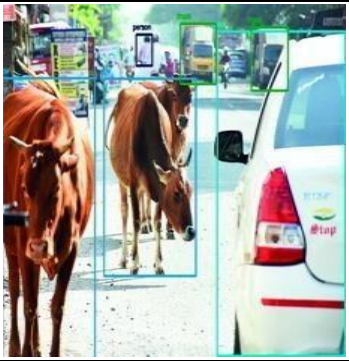
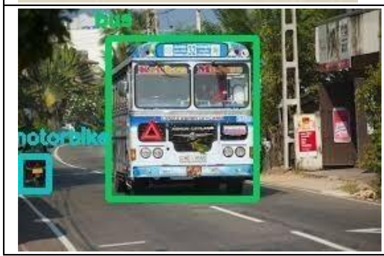
	one chair to the left onebed to the right
	one chair to the center onetv monitor to the right onevase to the right
	two cows to the left two trucks to the right one carto the right one person to the left
	one motorbike to the rightone car to the right
	three cars to the left fivecars to the right
	one bus to the center one motorbike to the left

Table 2.Real time indoor and outdoor test images and their corresponding output containing directions.

The following table shows the video frames captured in real time and produces a caption giving alerts to the VI of the approaching object if any based on the predefined, adjustable threshold

Video Frames	**Output**
	A person is approachin g towards you.

Table 3.Video frames of real time data and its corresponding output

B. *Analysis*

1) *Underlying Object Detection Algorithms*

Training dataset composed of images pertaining to but not limited to following classes- person, bus, truck, car, traffic light, cat, dog, sofa, chair, bench, bicycle, motorcycle, bed, desk, chair, sofa, vase etc. were used.Total images used for training were 50 thousand.

We trained different object detection models like RCNN, fast RCNN, faster RCNN, YOLO and SSD on this training dataset. In order to detect objects in real time, we tested our system to use different object detection algorithms and compared their accuracy and speed on a training dataset which consisted of images pertaining to different scenarios(indoor and outdoor). Total test images used were 10 thousand.

The following graphs depict the accuracy and speed of different algorithms.

Based on the analysis, YOLOV5 proved to be more accurate and fast as compared to other algorithms and hence YOLOV5 was used as the base for object detection for our system.

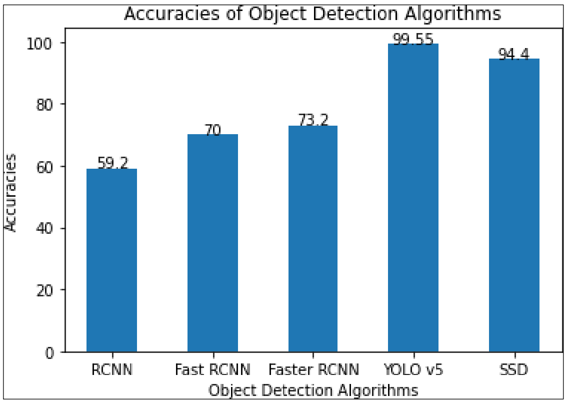


Figure 7.Graph depicting the accuracies of RCNN, Fast RCNN, Faster RCNN, YOLO v5 and SSD

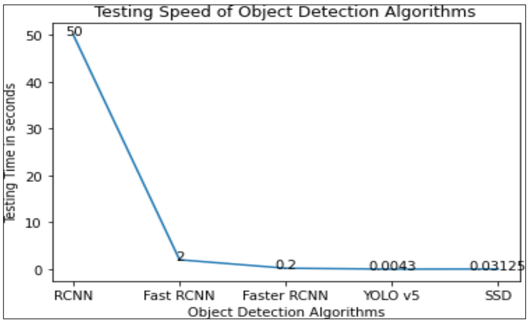


Figure 8.Graph depicting the Testing Speeds of RCNN, Fast RCNN, Faster RCNN, YOLO v5 and SSD

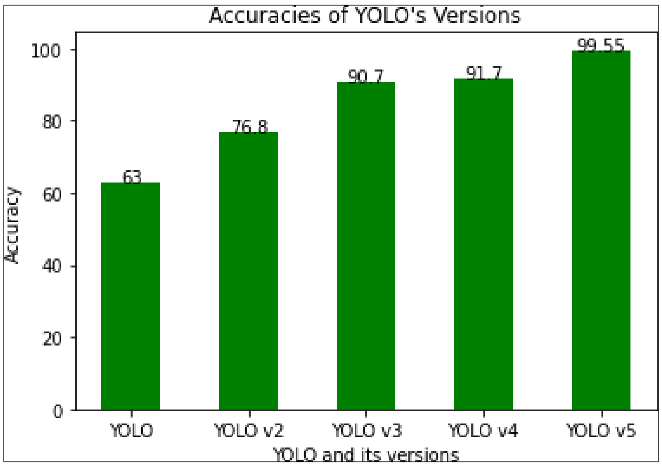


Figure 9.Graph depicting the accuracies of YOLO and its versions

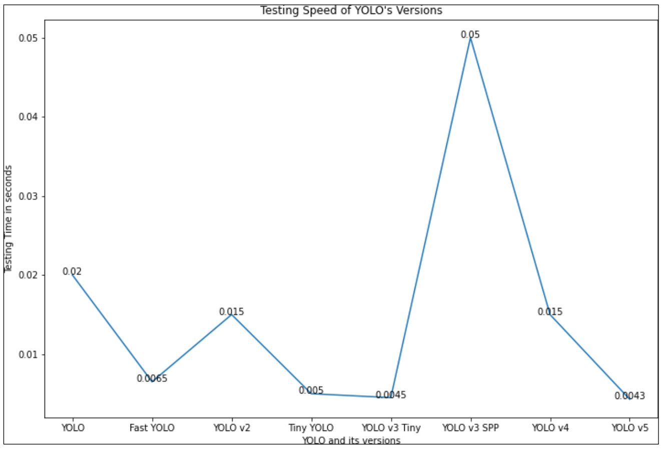


Figure 10.Graph depicting the accuracies of various versions of YOLO

2) Mobile Application Related Parameters

The following table shows a comparison between existing mobile apps and our mobile application VISION with respect to storage, data usage, permissions required, cache memory used.

Apps	**Storage**	**Data Usage**	**Permissions**	**Cache**
TapTapSee	16.36 MB	1.8 MB	Camera, Filesand Media, Location, Microphone	207 MB
Lookout byGoogle	156 MB	0 B	Camera, Filesand Media, Contacts	137 MB
Supersense	188 MB	0 B	Camera, Filesand Media, Location, Microphone	128 MB
Be My Eyes	52.63 MB	0 B	Camera, Microphone	349 MB
VISION	152 MB	0 B	Camera	175 MB

Table 4.Comparison of our app and the existing ones with respect to storage,data usage, permissions required, cache memory

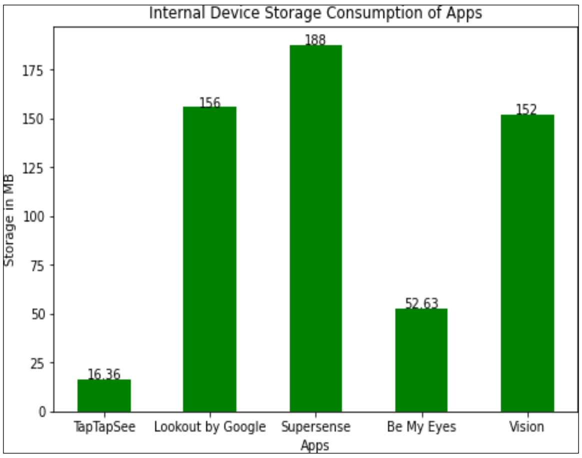


Figure 11.Bar graph depicting the device storage required for various apps

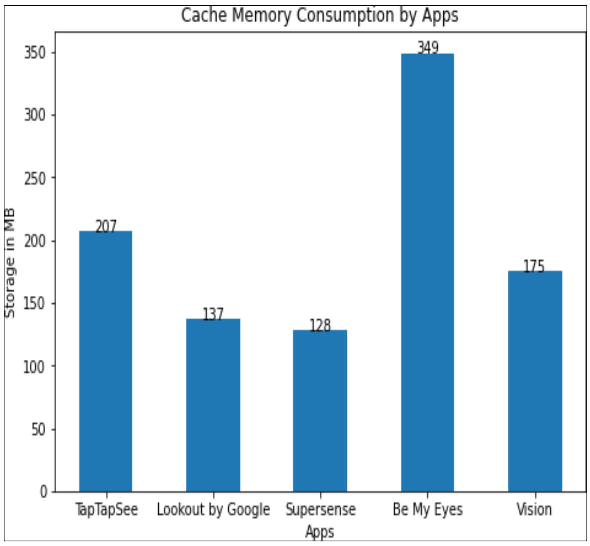


Figure 12.Bar graph depicting the Cache Memory consumed by various apps

3) Comparison of User Interface of Existing systems *with our mobile application VISION*

The structure of User Interface is very important with respect to Visually Impaired people. Our UI would prove more friendly for Visually Impaired People. As it supports tap based approach, Visually Impaired People would not need to press any button. As compared to other similar applications our app is more user friendly with respect to Visually Impaired People and designed in such a way that it makes life of Visually impaired people much more simpler and independent.

The following table depicts a comparison of our system's UI with other existing systems.

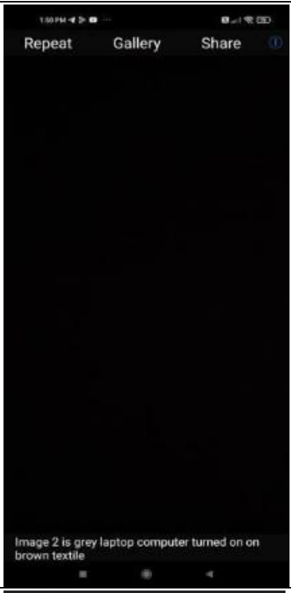
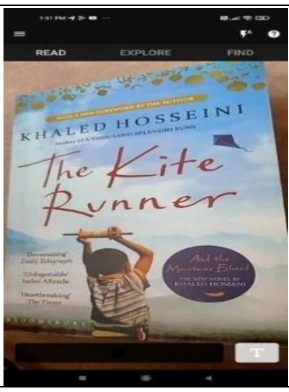
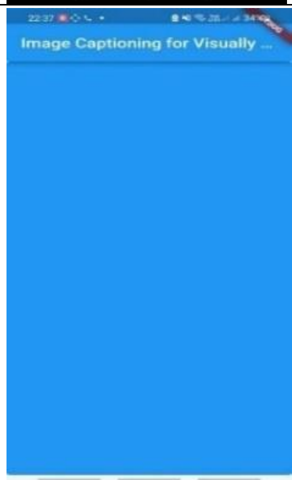
Apps	UI	Comments
TapTapSee		Describes an image given from the gallery.
Lookout byGoogle		Explores the surroundings and detects objects present. Read text and documents. Detects currency notes. Describes images given from gallery orphotos. Gives food labels.
Supersense		Read text. Detects objects from surroundings using video data. Finds the desired object from one's surroundings
		.
Be My Eyes		Pairs a VI with a volunteer.
VISION		User-friendly

Table 5.Comparison of existing systems with ours with respect to their User Interfaces (UI)

The User interface of our mobile application is designed in such a way that it caters to most of the difficulties of VI persons. The application can be easily opened using Google Assistant. When the application opens, audio instruction "Welcome to this App" is given. So that the person understands that the app has been opened. Then, in order to use features of the app, a person is instructed in audio format to tap anywhere in the bottom right of the screen to open the Camera. When the camera is opened, he is again notified that the camera has been opened and he needs to click anywhere on the screen to capture the image. Again when the image has been clicked he is notified about it . Finally in order to give direction the person is instructed to again click anywhere on the screen.

Thus our application helps to overcome the difficulty of the VI Person to click a particular button on a screen and thus is user friendly.

V. Conclusion

This paper provides a simple yet effective approach to aid VI People and make them independent and self-reliable. Approach of using coordinate geometry which is the base for providing directions to the user is different, accurate and light - weight as compared to the methods used in existing systems for giving directions.

The system is tested and verified on real time scenarios in order to understand its effectiveness as well as accuracy. Experiments related to object detection algorithms which isthe base of the whole system are also done in order to givebest outcomes.

Thus, an efficient system better than the existing ones is deployed in the form of mobile application that further cuts down the additional buying price of software integrated hardware systems with equivalent results. Security, privacy and gallery storage is also taken care of by the system.

VI. Future Scope

In the future, we intend to give directions based on margin-based image frames.

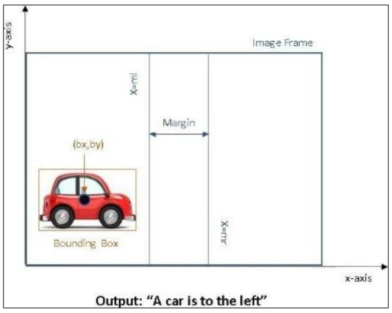


Figure 13.An image frame under computation for giving directions based on margin-based image frames

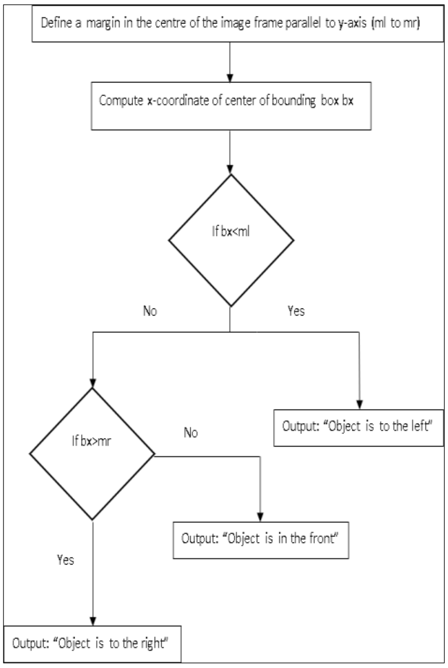


Figure 14.Flowchart for giving directions based on margin-based image frames

The object whether it is present to the center is very much subjective to frame of reference. Presently even if the center point of the object of interest is present slightly to the left or right of the center of the whole image then the algorithm detects it as object to the left or object to the right. Instead we wish to add a margin so that the image is detected to the left or right only if the difference between the center point of the object detected and center point of the image frame is much higher. In future, we wish to incorporate our video processing module in the mobile application.

References

[1] H. Jabnoun, F. Benzarti and H. Amiri, "Object detection and identification for blind people in video scene," 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), 2015

[2] Wong, Yan Chiew, et al. "Convolutional neural network for object detection system for blind people." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 11.2 (2019)

[3] Dunai, Larisa, et al. "3D CMOS sensor based acoustic object detection and navigation system for blind people." *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*. IEEE, 2012.

[4] Wojnowska-Heciak, Magdalena, et al. "Urban Parks as Perceived by City Residents with Mobility Difficulties: A Qualitative Study with In-Depth Interviews." *International journal of environmental research and public health* 19.4 (2022): 2018.

[5] Masud, Usman, et al. "Smart assistive system for visuallyimpaired people obstruction avoidance through object detection and classification." *IEEE Access* 10 (2022): 13428-13441.

[6] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: CVPR, 2005.

[7] D. G. Lowe, Object recognition from local scale-invariantfeatures, in: ICCV, 1999.

[8] Xiongwei Wu, Doyen Sahoo, Steven C.H. Hoi, Recent Advances in Deep Learning for Object Detection, Neurocomputing (2020)- Elsevier

[9] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, in: Proceedings of the IEEE, 1998.

[10] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: CVPR, 2014.

[11] Diwan, T., Anirudh, G. & Tembhurne, J.V. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* (2022).

[12] Yuvaraj Munian, Antonio Martinez-Molina, Dimitrios Miserlis, Hermilo Hernandez & Miltiadis Alamaniotis (2022) Intelligent System Utilizing HOG and CNN for Thermal Image-Based Detection of Wild Animals in Nocturnal Periods for Vehicle Safety, Applied Artificial Intelligence, 36:1, DOI:

[13] Jinjuan Wang, Xiliang Zeng, Shan Duan, Qun Zhou, Hao Peng, "Image Target Recognition Based on Improved Convolutional Neural Network", *Mathematical Problems in Engineering*, vol. 2022, Article ID 2213295, 11 pages, 2022.

[14] Nguyen, Van-Cam, Hong-Tuan-Dinh Le, and Huu-Thuan Huynh. "Hardware System Implementation for Human Detection using HOG and SVM Algorithm." *arXiv preprint arXiv:2205.02689* (2022).

[15] Bing Ou, Jingjing Yang, Wei Wang, "Analysis of People Flow Image Detection System Based on Computer Vision Sensor", *Journal of Sensors*, vol. 2022, Article ID 8099876, 7pages, 2022

[16] Başa, Berkant. "Implementation of Hog Edge Detection Algorithm Onfpga's." *Procedia-Social and Behavioral Sciences* 174 (2015)

[17] Yao Lu, An-An Liu, Yu-Ting Su, "Chapter 6 - Mitosisdetection in biomedical images", In Computer Vision and Pattern Recognition, Computer Vision for Microscopy ImageAnalysis, Academic Press, 2021.

[18] Zhu Mei, Yeu Wang, ”Research on Moving Target Detection and Tracking Technology in Sports Video Based on SIFT Algorithm”, Advances in Multimedia, vol. 2022, Article ID 2743696, 12 pages, 2022.
[19] Leqi Jiang, Zhihua Xia, Xingming Sun,“Chapter Three -Review on privacy-preserving data comparison protocols in cloud computing”, Advances in Computers, Elsevier, Volume 120, 2021.
[20] Satendra Pal Singh, Gaurav Bhatnagar, “Chapter 1 - Perceptual hashing-based novel security framework for medical images”, In Intelligent Data-Centric Systems, Intelligent Data Security Solutions for e-Health Applications, Academic Press, 2020.
[21] Prashanth Thinakaran, Diana Guttman, Mahmut Taylan Kandemir, Meenakshi Arunachalam, Rahul Khanna, Praveen Yedlapalli, Narayan Ranganathan, “Chapter 11 - Visual Search Optimization”, High Performance Parallelism Pearls, Morgan Kaufmann, 2015.
[22] Yang Song, Weidong Cai, “Chapter 4 - Visual feature representation in microscopy image classification”, In Computer Vision and Pattern Recognition, Computer Vision for Microscopy Image Analysis, Academic Press, 2021
[23] Pinar Muyan-Özçelik, Vladimir Glavtchev, Jeffrey M. Ota, John D. Owens, “Chapter 32 - Real-Time Speed-Limit-Sign Recognition on an Embedded System Using a GPU”, In Applications of GPU Computing Series, GPU Computing Gems Emerald Edition, Morgan Kaufmann,2011.
[24] Bappy, Jawadul H., and Amit K. Roy-Chowdhury. "CNNbased region proposals for efficient object detection." *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.
[25] Anamika Dhillon, Gyanendra K. Verma, Convolutional neural network: a review of models, methodologies and applications to object detection-Springer 2019.
[26] Youzi Xiao ,Zhiqiang Tian,Jiachen Yu1,Yinshu Zhang1 Shuai Liu1,ShaoyiDu2,Xuguang Lan2 ,A review of object detection based on deep learning-Springer 2019.
[27] H. Harzallah, F. Jurie, C. Schmid, Combining efficient object localization and image classification, in: ICCV, 2009
[28] R. Girshick, Fast r-cnn, in: ICCV, 2015.
[29] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: NeurIPS, 2015.
[30] Cong Tang 1,2,3, Yunsong Feng1,2,3, Xing Yang1,2,3, Chao Zheng1,2,3, Yuanpu Zhou1,2,3, The Object Detection Based on Deep Learning- IEEE 2017
[31] Hossain, MD Zakir, Ferdous Sohel, Mohd Fairuz Shiratuddin, and Hamid Laga. "A comprehensive survey of deep learning for image captioning." ACM Computing Surveys (CsUR) 51, no. 6 (2019)
[32] Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and tell: Lessons learned from the 2015 mscoco image captioning challenge." IEEE transactions on pattern analysis and machine intelligence 39, no. 4 (2016)
[33] Wu, Huafeng, et al. "Ship Fire Detection Based on an Improved YOLO Algorithm with a Lightweight Convolutional Neural Network Model." *Sensors* 22.19 (2022): 7420.
[34] Millos Awad et al “Intelligent eye:A mobile Application for assisting blind People”-2018 IEEE Middle East and NorthAfrica Communications Conference
[35] Rucha Doiphode et al “Be My Eyes: Android Voice Application for Visually Impaired People”
[36] Prajakta S, Sanika W, Anagha K “Object Detection: Literature Review” SoCPaR 2022

Author Biographies

Prajakta Saraf received her Bachelor of Technology degree in the field Information Technology from MKSSS’s Cummins College of Engineering for Women. Currently she is a working professional at Citi. Her research interests include Machine Learning (ML), Deep Learning, NaturalLanguage Processing (NLP), applications of ML models in various domains including Cybersecurity.

Sanika Watve received her Bachelor of Technology degree in the field Information Technology from MKSSS’s Cummins College of Engineering for Women. Currently she is a working professional at Citi. Her research interests include Machine Learning (ML), Deep Learning, NaturalLanguage Processing (NLP), applications of ML models in various domains including Cybersecurity.

Dr Anagha Kulkarni has been working with Cummins College of Engineering for Women, Pune for last 18 years. Her total experience is 28 years; She has published 16research papers and 1 book chapter. She has delivered keynote speeches, expert sessions in STTPs, FDPs and in Conferences. She has been a reviewer and session chair in many Conferences.