Object Detection and Localization for Visually Impaired

Dileep Reddy Bolla[1] *Hrishita Rauniyar* [2], *Sowmya DN* [3], *Rakshitha.T.P* [4], *Nabi Rasool* [5]

[1] Department of CSE, Nitte Meenakshi Institute of technology, Bangalore, Karnataka, India
[2,3,4] Department of CSE, Nitte Meenakshi Institute of technology, Bangalore, Karnataka, India

dileep.bolla@gmail.com, hrishitarauniyar123.np@gmail.com, sowmyadnrk@gmail.com, 1nt19cs151rakshitha@nmit.ac.in, nabirasoolshaik43@gmail.com

Abstract. Now a days Both temporary and permanent disabilities affect a large number of people in this one of the disabilities is blindness there are a lot of blind persons in the world. Nearly 390 lakh individuals are fully blind, and another 2850 lakh are purblind, or visually impaired, according to the World Health Organisation (WHO). Many supporting or guiding systems have been established, and are still being developed, to improve people's daily lives as they move from one place to another. Therefore, the main concept behind our suggested method is to provide an auto-assistance system for people who are visually impaired. As a result of their inability to see the object, the disabled person may find this auto-assistance system useful. To create an assisting system for blind persons, numerous methods have been put into place. Some systems are still being studied. The implemented models had a number of drawbacks when it came to object detection. We suggest a new method that uses CNNs (Convolutional Neural Networks) to aid those who are blind or visually impaired.

Keywords— CNN (Convolution Neural Network), YOLO, object detection, localization, Deep Learning

1. INTRODUCTION

According to the World Health Organisation (WHO), there are 285 million people worldwide who are blind or visually impaired. 39 million of them are blind [1]. The main conditions that cause visual impairments are refractive error, glaucoma, trachoma, corneal opacities, cataracts, diabetic retinopathy, and untreated presbyopia. [2]. People who are visually impaired (VIPs) have trouble performing activities of daily living (ADLs), such as finding common objects (indoors or outdoors) on their own or even with some help. They also have trouble moving around and interacting with their surroundings. The main challenges faced by VIPs are object detection and recognition, money identification, textual information (signs, symbols), translation, mobility/navigation, and safety[3].Several methods, platforms, tools, and software have been created in the assistive technology field in the past to help VIPs carry out tasks that they were formerly unable to do [4]. These

solutions often consist of electronic gadgets with cameras, sensors, and microprocessors that can decide and give the user input via touch or sound. Although many of the currently available object detection and recognition systems make claims of great accuracy, they are unable to deliver the data and qualities required for tracking VIPs and ensuring their safe movement [5] Even if blind persons are unable to see objects in their environment, learning about them is still beneficial. Additionally, a tracking system must be created so that VIPs' families can keep tabs on their whereabouts.

In light of the aforementioned requirements, this study offers a smart system that executes real-time object localisation and recognition. The user receives audio feedback as soon as the system locates the object. Following the user's identification of a well-known object, like an automobile, they will hear the word "car". The user's location and a screenshot of the most recent scene they watched are also regularly stored on a server that family members can access via an app to track the user. Since Mobile-Net architecture has a low level of computational complexity and can run on low power end devices, it is used for object detection and recognition. Since wearable hardware resources are limited and the system's feedback regarding the object's name needs to be as accurate as possible, complex, cutting-edge object recognition methods might not be practicable as the principal strategies.

1.1 The proposed work's objectives:

The main objective of this proposal is to develop a novel system with the properties listed below. First, for instantaneous object recognition and identification, a deep learning architecture is utilised.. It says the names of things that are visible to the camera's sight, or those that are in the current frame. It periodically notifies a web server of the user's location. The webserver receives a live stream and snapshots from it. Important family members can use a web-based interface to track their location when they're at home. It is a feature that the user may choose to employ that ensures their security and privacy..

2. Related work

More than one-fourth of the 36 million drowsy people worldwide reside in India. One of the most difficult problems dazzle schools face today is teaching the dazzle how to avoid unemployment among their population.Despite the fact that many schools use Braille to eliminate the lack of education among their students, its steep learning curve, poor accessibility, and high cost makes it exceedingly unapproachable. Less than 10% of the 12 million blind people in India, according to Braille education measurements, are able to read Braille. Relevant work

2.1 Assistive Innovation

To solve this problem, a framework that can aid the visually impaired in reading needs to be developed. Therefore, the suggested solution is to design a low-cost wearable device that uses computer vision to analyse any shape of content surrounding the client in various configurations and lighting circumstances. The system utilizes a Raspberry Pi and a compatible camera to capture the information surrounding the visually impaired or dazed person and read it to them in their native language. As the device locates various things, a sensor is also integrated to alert the user of the distance to the nearest protest at his eye level. The system is constructed using a combination of photo processing, machine learning, and discourse fusion techniques. The calculated watched precision for both the question acknowledgment computations and the optical character acknowledgment calculations was determined to be 84%.

2.2 Smart Specs:

The World Health Organisation estimates that out of a global population of 7.4 billion, 285 million people are considered to be physically impaired. It is watched that they are still finding it troublesome to roll their day nowadays life and it is imperative to require a vital degree with the developing advances to assist them to live the current world independent of their disabilities. Within the rationale of supporting them, We have proposed a savvy spec for the blind persons which can perform content discovery in this manner deliver a voice yield. This will offer assistance the outwardly impeded people to study any printed content in vocal frame. A spec inbuilt camera is utilized to capture the content picture from the printed content and the captured picture is analyzed utilizing Tesseract-Optical Character acknowledgment (OCR). The recognized content is at that point changed over into discourse employing a compact open-source program discourse synthesizer, Talk. At long last, the synthesized discourse is created by the earphone by TTS strategy. In this project Raspberry Pi is the most target for the execution, because it gives an interface between camera, sensors, and picture preparing comes about, whereas moreover performing capacities to control fringe units (Console, USB etc.,).

2.3 Assisted Development

Numerous endeavors have been contributed within the final a long time, based on brilliant gadgets and data innovation, in arrange to create Estimated time of arrival supplies as a substitute for the misplaced locations of dazzle and outwardly impeded people. As a result, there are suitable arrangements for a few of the problems in this region. Within the final a long time, the conventional apparatuses utilized by outwardly impeded to explore in genuine open air situations (white can and directing pooches) are to be substituted with electronic travel helps (Estimated time of arrival) These gadgets, based on sensor innovation and flag preparing, are able to move forward the portability of dazzle clients in obscure or powerfully changing environment. Within the display paper, the foremost imperative hypothetical and common sense comes about gotten within the field of ETAs are displayed to begin with. A few unique comes about of the author's group, which incorporate modern concepts in this field, like coordinates environment for helped development, acoustical virtual reality (AVR), bioinspired arrangements are at that point examined in more detail.

2.4 CNN Based Relationship Calculation to Help Outwardly Disabled Persons

A CNN based relationship [4] calculation to help outwardly disabled individuals is clarified. Notwithstanding of the variant proposed, contain a visual handling unit within the structure of the frameworks that helps individuals with visual impedances is reckless essential, given the large number of data that can be extricated from pictures procured. This paper presents a relationship calculation based on the utilization of cellular neural systems (CNNs) that can progress the highlights of assisting systems, to provide more data from the environment to outwardly impeded people. The foremost of operations (calculations) included within the proposed calculation is achievable by parallel preparing. In this way, it can diminish the computing time and the computing time will not increment relatively with expanding the estimate of the format images.

2.5 Multicore Convenient Framework for Helping Outwardly Impeded People

The convenient framework [5] is constructed around a smartphone but also uses tactile modules. It covers indoor and open-air developments of outwardly disabled individuals. Tests have shown that the framework's proficiency can be improved with the advancement of Android-based versatile gadgets.

This paper presents a convenient framework to help outwardly impeded individuals in indoor and open-air situations. Its employments diverse sensors to identify impediments and direct them in their development with the help of GPS and compass. The most portion of the framework comprises of a multicore Android smartphone. Other tactile modules detect impediments and communicate pertinent data to the most portion. For separate monitoring, the framework can communicate remotely.

2.6 Smart Glasses for the Outwardly Impeded People

People with visual disability confront different issues in their standard of living [6] as the cutting edge assistive gadgets are regularly not assembly the shopper necessities in term of cost and level of help. This paper presents an unused plan of assistive keen glasses for outwardly disabled understudies. The objective is to help with different day by day errands utilizing the advantage of the wearable plan arrangement. As a verification of the concept, this paper as it were presents one illustration application, i.e. content acknowledgment innovation that can offer assistance perusing hardcopy materials. The building taken a toll is kept moo by utilizing single board computer raspberry pi 2 as the heart of preparing and the raspberry pi 2 camera for picture capturing. Explore comes about illustrate that the model is working as intended.

2.7 A Savvy Wearable Route Framework for Outwardly Impaired

Smart gadgets [7] are getting to be more common in our everyday lives; they are being joined in buildings, houses, cars, and open places. Besides, this innovative transformation, known as the Web of Things (IoT), brings us unused openings. A assortment of route frameworks has been created to help daze individuals. However, none of these frameworks are associated to the IoT. The objective of this paper is to execute a moo fetched and moo control IoT route framework for dazzle individuals. The framework comprises of an cluster of ultrasonic sensors that are mounted on a abdomen belt to overview the scene, iBeacons to recognize the area, and a Raspberry Pi to do the information handling. The Raspberry Pi employments the ultrasonic sensors to distinguish the deterrents, and give sound prompts through a Bluetooth headset to the client.

3.Design and Development of Object Detection

In the proposed work, a digitally altered image is stored in a frame buffer, which is a matrix of pixels with W columns and H rows. Let $(0x, 0y)$ be the focal point of the lens in the frame coordinates, (i, j) be the discrete frame coordinates of the picture origin in the upper left corner, and (x, y) be the image coordinates as in Fig.1 [15].

$$x = (i - 0x) \times w \tag{1}$$

$$y = (j - 0x) \times H \tag{2}$$

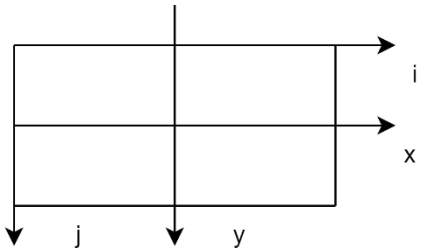

Fig. 1: Mapping frame coordinates to image

3.1 depth estimation and calculation:

The triangle trigonometry model that forms the basis for the depth calculation is shown in Fig. 2. The bounding boxes from the object detection component are supplied as input to the depth computation in the following format: $(ymin, xmin, ymax, xmax)$. The depth of the objects is determined by comparing the centres of these bounding boxes in the left and right frames. The boxes' centres should have the following shapes: $(box1\_xc, box1\_yc)$ *and* $(box\_xc, box2\_yc)$.

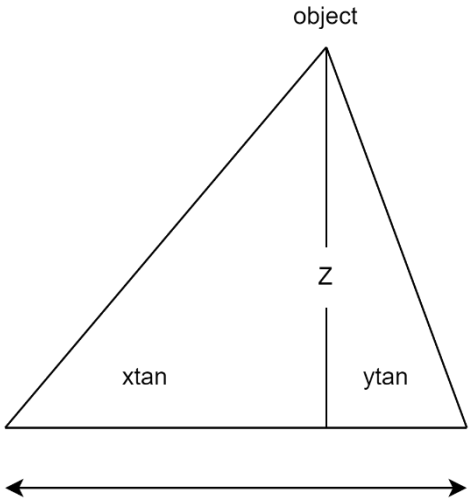

Fig. 2: Triangulation Trigonometry

Fig. 3 illustrates the calculation of camera angles in the vertical and horizontal planes in 2D. The tone of the camera's angle_width and angle_height is used to calculate the vertical distance from the centre of the picture, which serves as a reference point for the origin as in Fig.3. Let these distances be represented by x_adj *and* y_adj.

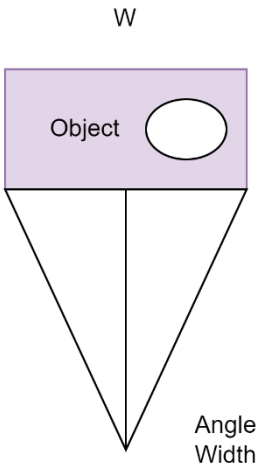

Fig.3 Camera Angle calculation

The angle formed by them and the box centres can then be determined as follows:

$$xtan = xxadj \tag{3}$$

$$ytan = yyadj \tag{4}$$

Using equations 3, 4, and the distance between the cameras (p), the depth of the centres of the boxes is finally determined as follows:

$$Z = p1xtan+1\text{ytan} \tag{5}$$

3.2 YOLO Model:

Instead than seeing object detection as a classification problem, YOLO views it as a single regression problem. The name "YOLO" refers to the fact that this system only looks at the image once to identify the items and their locations The system creates a $S\ x\ S$ grid from the image. Each of these grid cells displays B bounding boxes and the confidence scores related to them. The model's level of confidence that the box contains an object can be seen in the confidence score, which also indicates how reliable the model thinks the box is in making predictions. Equation 6 can be used to get the confidence score.

$$C = pr(object) * IoU \tag{6}$$

IoU stands for the intersection over union of the actual data and the anticipated box. If a cell doesn't contain any objects, the confidence score for that cell should be zero.

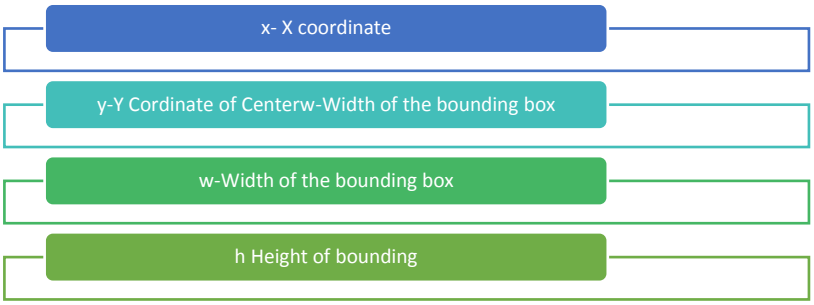

Fig.4. Bounding Box Predictions

Each bounding box is composed of five predictions: x, y, w, h, and confidence, where (x, y) are the coordinates of the box's centre. These coordinates are determined in reference to the grid cell boundaries. w: The width of the bounding box. h: The height of the bounding box as in Fig.4. Additionally, for C conditional class probabilities, each grid cell projects $Pr(Class\ i|Object)$.

$$Pr(class\ i|Object) * \Pr(Object) * IoU = \Pr(Class\ i) * IoU \quad (7)$$

The final predictions are encoded as an

$$S = S \times (B * 5 + C) \quad (8)$$

It only estimates one set of class probabilities per grid cell, regardless of the quantity of boxes *B.* These conditional class probabilities are multiplied by the individual box confidence forecasts during testing to generate class-specific confidence ratings for each box as in Fig.5. Both the likelihood of that class and how well the box fits the object are shown in these evaluations.

4. Proposed system

The brain of our proposed paradigm is the Raspberry Pi. Because we want the results to be inside a sound frame, We've chosen to deploy a speaker. Additionally, the Raspberry Pi is compatible with high-bass headphones. We are employing the Raspberry Pi (3 B+) method.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig.5. Network Architecture of YOLO

In order to provide clients with flexibility, we chose to employ a control bank as the source of the Raspberry Pi's power supply.. Its use is based on the Raspberry Pi, one of the most well-liked single-board computers. The Raspberry Pi's OpenCV software makes it easy to perform all the necessary calculations and operations for picture processing. We are utilising a 32 GB lesson 10 SD card for our Raspberry Pi. Additionally, we are using a USB camera in place of the Raspberry Pi camera because its wiring is stiff and challenging to maintain. Pi (Rpi). We demonstrate a Rpi-based YOLO calculation as in Fig.5. A speaker is connected to one of the Raspberry Pi's USB ports to serve as a simple speaking device. Because we require portability, we are employing the 5 volt control bank as a control supply

The graphic below displays the block diagram of our system as in Fig.6 which consists of a camera, raspberry pi, speaker, and control bank. Our framework tool's first requirement is to secure images, and this is taken care of by the USB camera connected to the Raspberry Pi's USB port.

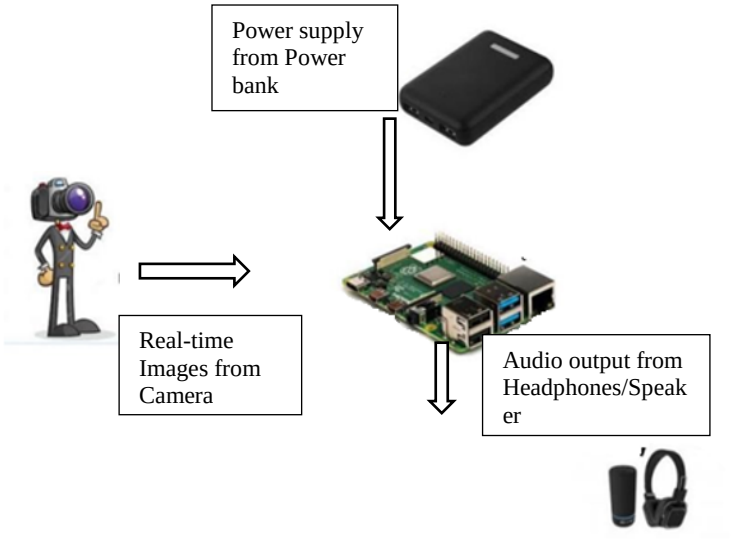

Fig. 6. Block Diagram of User Side

4.1 Implementation of the Proposed model

A communication diagram that illustrates how and in what order various items interact with one another may be included in the flowchart as in Fig.7. The first stream of our system, in which the client begins and dons the frame, is described by the flowchart above. As soon as the Raspberry Pi (Rpi) is turned on, its internal code or process will start running. Until the Raspberry Pi is turned on, the code never ends. Before browsing the content record containing details on the titles of the lessons, YOLO weights, and arrangement records, Rpi will first import all the required libraries, including OpenCV, Pyttsx3, Time, and NumPy. The code will then start the correct camera.

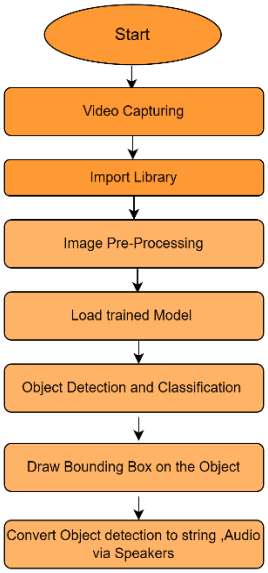

Fig 7: Flowchart of system

Once the camera records real-time outlines at a rate of one frame per second (fps), the code will examine the incoming image/frame and modify the width and height to an acceptable level. Then, using YOLO as an example, this changed form is linked to the protest location algorithm. Before sending this modified image to the YOLO weights and YOLO setup records, a 'BLOB from image' is built. The OpenCV function blob From Image was used to obtain (modify) expectations from image categorization. In order to provide us with our bounding boxes, course ids, and related lesson probabilities, the code then does a forward pass of the YOLO object detector. Besides being rapid, another benefit of YOLO is that it offers three ways to advance its execution:

4.2 Intersection over Union (IoU)

decides which predicted box is giving a good outcome. It calculates the IoU of the actual bounding box and the predicted bounding box.

4.3 Anchor Boxes and Non-max Suppression

The proposed work suppresses weak, overlapping bounding boxes and recognises several items in a single grid. Encourage, which impacts where the objects are situated, divides the outlines into a 3x3 framework. For people with obvious limitations, our design strives to provide an auditory output. The discovered item labels are transformed into discourse using the pyttsx3 module.

Last but not least, the framework **as in Fig.8** will provide discourse after properly acknowledging a protest and in accordance with grids, expressing the name of the protest along with its lattice title , for example, "Mid cleared out car" or "Mid right car." Making a difference in how people with external impairments see the items in their field of vision.

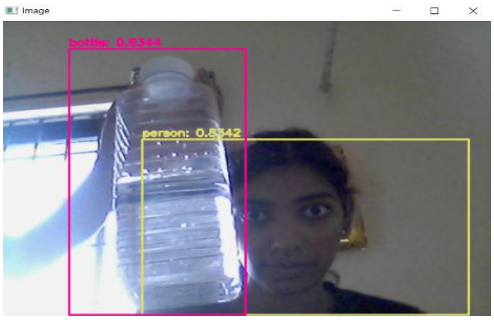

Fig 8. Division of Image into grids

5. RESULT AND CONCLUSION

Framework based helping arrange has been proposed in arrange to help the purblind individuals and totally daze individuals. The profound learning based methods conducted by testing utilizing Open CV and keras has formed a fruitful strategy that's multiscale and valuable for the applications utilized interior the environment.

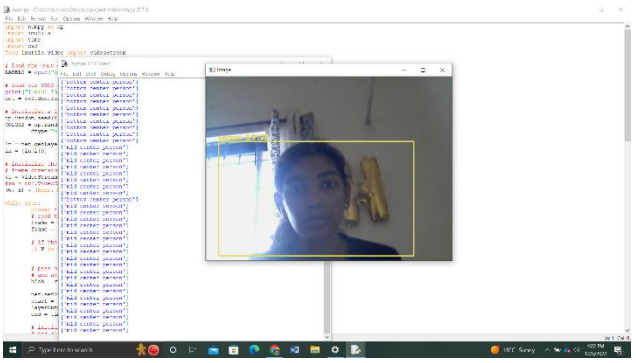

Fig 9. Person Recognized

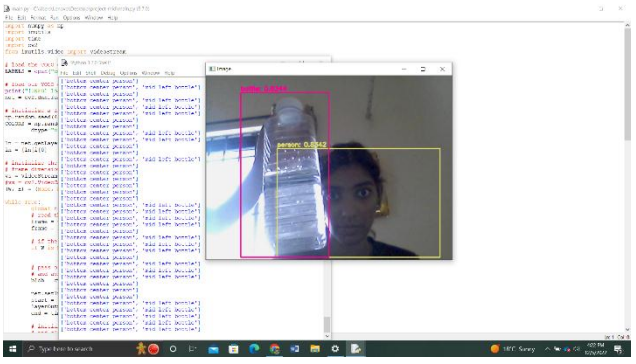

Fig 10. Bottle and Person Recognized

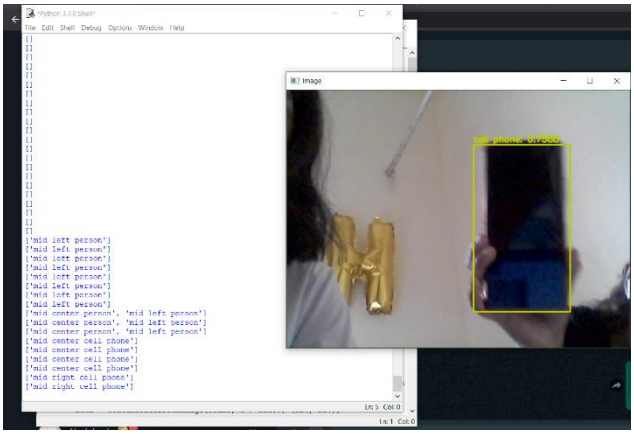

Fig 11. Cell Phone Recognized

In the proposed work we are able to recognize the person as in Fig.9 with a boundary box highlighted. Further double image with bottle and person are highlighted as in Fig.10, finally we are able to train and test with objects say mobile phone is also succesfully detected in Fig.11.

Table 1. Testing accuracy of proposed system

Methods	**Testing Accuracy**
CNN	95%
YOLO v3	95.12%

Based on the research work carried out the model is tested and validated using CNN and Yolo V3 and the accuracy is recorded as in Table 1. For the proposed work YOLO v3 performes better when comared to CNN model.

Acknowledgment

We thank Nitte Meenakshi Institute of Technology's Centre of Multidisciplinary research in Cyber Security and IoT for providing the necessary infrastructure and technical support.

References

1. M.P. Arakeri, N.S. Keerthana, M. Madhura, A. Sankar, T. Munnavar, “Assistive Innovation for the Outwardly Impeded Utilizing Computer Vision”, Universal Conference on Propels in Computing, Communications and Informatics (ICACCI), Bangalore, India, pp. 1725-1730, sept. 2018.
2. R. Ani, E. Maria, J.J. Joyce, V. Sakkaravarthy, M.A. Raja, “Smart Specs: Voice Helped Content Perusing framework for Outwardly Impeded People Utilizing TTS Method”, IEEE Universal Conference on Developments in Green Vitality and Healthcare Advances (IGEHT), Coimbatore, India, Damage. 2017.
3. V. Tiponuţ, D. Ianchis, Z. Haraszy, “Assisted Development of Outwardly Impeded in Open air Environments”, Procedures of the WSEAS Worldwide Conference on Frameworks, Rodos, Greece, pp.386-391, 2009.
4. L. Ţepelea, A. Gacsádi, I. Gavriluţ, V. Tiponuţ, “A CNN Based Relationship Calculation to Help Outwardly Disabled Persons”, IEEE Procedures of the Universal Symposium on Signals Circuits and Frameworks (ISSCS 2011), pp.169-172, Iasi, Romania,2011.

5. P. Szolgal L. Ţepelea, V. Tiponuţ, A. Gacsádi, "Multicore Convenient Framework for Helping Outwardly Disabled People", 14th Worldwide Workshop on Cellular Nanoscale Systems and their Applications, pp. 1-2, College of Notre Woman, USA, July 29-31, 2014.
6. E.A. Hassan, T.B. Tang, "Smart Glasses for the Outwardly Impeded People", 15th Universal Conference on Computers Making a difference Individuals with Uncommon Needs (ICCHP), pp. 579-582, Linz, Austria, 2016.
7. M. Trent, A. Abdelgawad, K. Yelamarthi, "A Savvy Wearable Route Framework for Outwardly Impaired", 2nd EAI worldwide Conference on Shrewd Objects and Advances for Social Great (GOODTECHS), pp. 333-341, Venice, Italy, 2016.
8. Jae Sung Cha, Dong Kyun Lim and Yong-Nyuo Shin, "Design and Usage of a Voice Based Route for Outwardly Disabled Persons", Universal Diary of Bio-Science and Bio-Technology, Vol. 5, No. 3, pp.61-68, June 2013.
9. S. Khade, Y.H. Dandawate, "Hardware Execution of Deterrent Location for Helping Outwardly Impeded Individuals in an New Environment by Utilizing Raspberry Pi", Keen Patterns In Data Innovation And Computer Communications, SMARTCOM 2016, vol. 628, pp. 889-895, Jaipur, India, 2016. [10] R. C. Gonzalez, R. E. Woods and S. L. Eddins, "Digital Picture Preparing utilizing MATLAB", Pearson Instruction, 2004.
10. D. R. Bolla, J. J. J, S. S. Palle, M. Penna, Keshavamurthy and Shivashankar, "An IoT Based Smart E-Fuel Stations Using ESP-32," 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2020, pp. 333-336, doi: 10.1109/RTEICT49044.2020.9315676.
11. S. shankar, J. J. Jijesh, D. R. Bolla, M. Penna, P. V. Sruthi and A. Gowthami, "Early Detection of Flood Monitoring and Alerting System to Save Human Lives," 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2020, pp. 353-357, doi: 10.1109/RTEICT49044.2020.9315556.
12. Bolla, D. R., et al. "Real-time data fusion applications in embedded sensor network using TATAS." Indian Journal of Science and Technology 10.13 (2017): 1-7.
13. D. R. Bolla, Shivashankar, R. Praneetha and B. S. Rashmi, "LI-FI TECHNOLOGY BASED AUDIO AND TEXT TRANSMISSION," 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2019, pp. 1532-1537, doi: 10.1109/RTEICT46194.2019.9016856.
14. D. R. Bolla and Shivashankar, "An efficient protocol for reducing channel interference and access delay in CRNs," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 2247-2251, doi: 10.1109/RTEICT.2017.8257000.
15. Bolla, D.R., Naidu, P.R., J J, J., T.R, V., Palle, S.S., Keshavamurthy (2023). Energy-Efficient Dynamic Source Routing in Wireless Sensor Networks. In: Shetty, N.R., Patnaik, L.M., Prasad, N.H. (eds) Emerging Research in Computing, Information, Communication and Applications. Lecture Notes in Electrical Engineering, vol 928. Springer, Singapore. https://doi.org/10.1007/978-981-19-5482-5_65.