ISSN No:-2456-2165

Vision Assistance System using Object Detection and Artificial Intelligence

[1]Swati Rai(20BCE0996); [2]Adnan Alam(20BCE2156); [3]Ish Gupta(20BCE2394);
[4]Dr. Narayanamoorthi M, (10454)
[1,2,3]Student; [4]Assistant Professor Grade 1
School of Computer Science, Vellore Institute of Technology
Vellore Institute of Technology, Katpadi, Vellore – 632014

Abstract:- Visual impairment, affecting 20% of India's population, poses challenges to independence and mobility. To address this issue, an Integrated Machine Learning System has been developed, enhancing autonomy and safety. It offers real-time object recognition, clear voice feedback, and unique distance calculations for safety. Designed with a user-centric approach, the system allows customization, supports offline functionality, and addresses limited internet access in certain regions. Continuous improvement and collaboration with the visually impaired community, along with robust data privacy measures, are integral to this initiative. This effort significantly improves the quality of life for visually impaired individuals in India, fostering inclusivity and independence. Beyond object recognition, the system's real- time distance calculations provide safety through proximity warnings, enhancing independence and safety for visually impaired individuals.

Keywords:- *Visually Impaired, Artifical Intelligence, Object Detection, Single Shot Detection*

I. INTRODUCTION

Assisted vision technologies, often referred to as assisted visual technology, constitute a captivating and profoundly influential domain that converges technology, healthcare, and accessibility. This field is driven by the remarkable potential it possesses to significantly enhance the lives of individuals with visual impairments and elevate their overall quality of life. At its core, assisted vision technologies seek to address a multitude of challenges faced by visually impaired individuals and open doors to a world where independence, inclusivity, and safety are not distant ideals but tangible realities. The significance of delving into this topic lies in the myriad motivations that underscore the transformative potential of assisted vision technologies.

II. MOTIVATION

This paper seeks to explore the following key motivations, which highlight the remarkable breadth and depth of impact this field can bring:

- Enhancing Independence: Assisted vision technologies empower individuals with visual impairments to navigate the world independently. These technologies offer real-time information about their surroundings, enabling them to perform daily tasks, travel, and engage in activities that may have otherwise seemed insurmountable.
- Breaking Down Barriers: Visual impairment can create barriers in education, employment, and social interactions. Assisted vision technologies act as a bridge, offering tools that grant access to information, digital content, and communication, enabling visually impaired individuals to fully participate in various aspects of life.
- Improving Safety: Navigating unfamiliar environments can be perilous for individuals with visual impairments. Assisted vision technologies provide vital capabilities, including hazard detection, obstacle avoidance, and pedestrian detection, which significantly enhance safety both indoors and outdoors.
- Education and Learning: Access to education is a fundamental right, yet it can be impeded by visual impairments. Assisted vision technologies facilitate learning by providing visually impaired students with access to textbooks, online resources, and interactive learning platforms, thereby promoting equal educational opportunities.
- Employment Opportunities: Meaningful employment can be elusive for people with visual impairments. Assisted vision technologies bridge this gap by offering tools that empower visually impaired individuals to access and contribute to digital workspaces, opening up new career possibilities.
- Innovation and Research: The exploration of assisted vision technologies offers a unique avenue for innovation in fields such as computer vision, artificial intelligence, and wearable technology. Researchers and engineers can collaborate to create novel solutions that push the boundaries of what is achievable in assistive technology.
- Human-Centered Design: The development of assisted vision technologies necessitates a human-centered design approach, encouraging engineers and designers to work closely with visually impaired individuals to create solutions that genuinely meet their needs, preferences, and aspirations.

The motivations outlined above underscore the profound impact and the far-reaching implications of assisted vision technologies. This paper aims to delve deeper into these motivations, exploring the current state of the field, emerging technologies, and the challenges that lie ahead. By doing so, we hope to contribute to a greater understanding of the transformative potential of assisted vision technologies and their role in creating a more inclusive and equitable society.

III. RELATED WORK

This paper [1] titled "FaceNet: A Unified Embedding for Face Recognition and Clustering," published in 2020 by Florian Schroff, Dmitry Kalenichenko, and James Philbin, explores the use of Convolutional Neural Networks (CNNs) in the context of face recognition and clustering. The authors describe their experimental setup, which involves training the CNN using Stochastic Gradient Descent (SGD) with standard backpropagation and AdaGrad. In the experiments conducted, the authors achieve significant results with this approach. They highlight the surprising effectiveness of their method. However, they acknowledge some limitations and suggest future directions for research. One limitation is the potential for diminishing returns in improving the v2 embedding over v1 while maintaining compatibility. The paper also discusses the possibility of training smaller networks suitable for mobile devices that remain compatible with larger server-side models.The findings of this paper contribute to the field of face recognition and clustering, opening doors for further research and potential improvements in the techniques and models used for this purpose.

This paper [2] titled "A Survey on Recent Advances in AI and Vision-Based Methods for Helping and Guiding Visually Impaired People," authored by Helene Walle, Cyril De Barthelemey erres, and Gilles Venturini in 2022, offers an overview of the latest advancements in AI and vision-based techniques aimed at assisting visually impaired individuals.The authors highlight the significant progress in AI technology, which has led to increased robustness and efficiency in helping visually impaired people. They suggest that future innovations hold the potential to provide safer and more effective wayfinding systems for BVIPs (Blind and Visually Impaired People).However, the paper acknowledges certain limitations associated with these technologies. Firstly, the acquisition devices required for vision- based solutions may come with a high cost, which could pose a barrier to widespread adoption. Additionally, adapting these AI models to specific conditions, such as varying environments and individual needs, can be challenging. This survey paper offers valuable insights into the evolving field of AI and vision-based methods for assisting the visually impaired, emphasizing the potential for further advancements while highlighting current limitations that need to be addressed.

This paper [3] titled "Vision-Based System for Assisting Blind People to Wander Unknown Environments in a Safe Way," authored by Andrés A. Díaz-Toro, Sixto E. Campaña-Bastidas, and Eduardo F. Caicedo-Bravo in 2021, explores the development of a vision-based system aimed at helping blind individuals navigate unfamiliar environments safely.The authors highlight the potential for transferring technology and insights from the field of autonomous cars to assistive tools for the visually impaired. They point out that both domains share similar requirements, including the need for real-time performance, the ability to operate in unknown and changing environments, and the necessity for safety. They emphasize the increasing accuracy and portability of 3D vision sensors, as well as the growing computational power and portability of embedded processors, which make such a transfer feasible. However, the paper acknowledges certain limitations of the technology. One notable limitation is the potential overloading of the sense of hearing, especially in dynamic environments, which can make it challenging for users to process auditory feedback effectively. Additionally, the use of tactile feedback devices like Braille displays, haptic belts, vests, and gloves may require a period of training for users to interpret commands adequately, particularly in dynamic settings.This paper offers insights into the potential for leveraging vision-based systems to assist blind individuals in navigating unfamiliar environments while also recognizing some of the challenges and limitations that need to be addressed in the development of such systems.

This paper [4] titled "Efficient Multi-Object Detection and Smart Navigation Using Artificial Intelligence for Visually Impaired People," authored by Rakesh Chandra Joshi, Saumya Yadav, Malay Kishore Dutta, and Carlos M. Travieso-Gonzalez in 2020, presents a technology designed to assist visually impaired individuals in detecting objects and navigating their surroundings using artificial intelligence.The authors report that the developed technology has proven to be highly useful for visually impaired users. It allows them to understand their surrounding environment effortlessly during navigation, eliminating the need for extensive manual effort. The system's ability to provide assistance without requiring prior knowledge about the position, shape, and size of objects and obstacles is considered a notable advantage.However, the paper also recognizes certain limitations of this technology. One significant limitation is the cost associated with the use of various distance sensors, which may render the system unaffordable for some users. To enhance the system's capabilities, the authors suggest the incorporation of additional sensors to detect objects and obstacles in different scenarios, such as stairs and other trajectories. This expansion of sensor usage could provide a wider range of assistance to visually impaired individuals.This paper highlights the efficiency of a multi-object detection and navigation system for the visually impaired, demonstrating its usefulness while acknowledging the need for cost-effective solutions and further sensor enhancements to address various navigation scenarios.

This paper [5] titled "Assisting Visually Impaired People by Computer Vision - A Smart Eye," authored by Arun Kumar Ravula, Pallepati Vasavi, and K. Ram Mohan Rao in 2021, discusses a computer vision-based system designed to assist visually impaired individuals.The paper highlights that the most important feature of this technology is distance estimation, as it is highly valuable for the user. When the user utilizes the device, it can detect objects placed in front of them and audibly relay information about the object's identity along with the distance of the object from the user. This feature provides essential assistance to visually impaired individuals in understanding and navigating their environment. However, the paper also acknowledges a limitation of the technology. While it excels in detecting and providing information about objects and distances, it does not support activities like playing games,

such as tic-tac-toe, using voice commands. This limitation underscores the specific focus of the system on assisting with object detection and navigation rather than entertainment or recreational activities.

The paper[6] titled "Vision-Based Navigation for Visually Impaired People: A Review" authored by Massimiliano Lippi, Saverio Iacoponi, Laura Burattini, and Luca Fanucci in 2020 presents a comprehensive review of the state of the art in vision- based navigation systems designed to aid visually impaired individuals. These systems aim to empower individuals with visual impairments to navigate their surroundings safely and independently. This review is a testament to the growing interest and importance of leveraging computer vision and related technologies to enhance the quality of life for visually impaired individuals. The authors delve into various aspects of vision-based navigation, addressing both indoor and outdoor scenarios. They explore the wide array of techniques and technologies that have been developed to aid individuals with visual impairments in their mobility and orientation. The central theme of the paper revolves around how these systems utilize computer vision methodologies, wearable devices, and innovative technologies to provide spatial awareness, detect obstacles, and assist with route planning.

This paper [7] titled "Face Detection and Recognition," authored by Jashanpreet Kaur, Akanksha, and Harjeet Singh in 2018, discusses the significant computer vision tasks of face detection and recognition, along with their wide-ranging applications, including security, surveillance, and access control. The paper explains that face detection involves identifying the presence of faces in an image or video, while face recognition entails determining the identity of an individual based on their facial features. Notably, there has been substantial progress in the development of face detection and recognition algorithms in recent years, primarily driven by advancements in machine learning, particularly deep learning. Deep learning algorithms have achieved state-of-the-art results on various face detection and recognition benchmarks. However, the paper also acknowledges certain limitations of these technologies. One of the most significant limitations is their sensitivity to factors like pose, lighting, and occlusion. Additionally, face detection and recognition algorithms may struggle to identify faces in crowded scenes or when faces are partially obscured by objects like glasses or masks. Another critical limitation is the potential for bias in these algorithms, which can result in reduced accuracy when identifying faces of people of color or women. Such bias often stems from the use of non-representative training datasets. This paper provides insights into the state of face detection and recognition, highlighting both the substantial progress made and the challenges and limitations that still need to be addressed for more robust and equitable performance.

The paper titled "Object Detection Based on the Improved Single Shot MultiBox Detector," authored by Songmin Jia, Chentao Diao, and Guoliang Zhang in 2019, presents notable enhancements to the Single Shot MultiBox Detector (SSD) object detection algorithm. These enhancements include the addition of a shallow object detection layer to improve small object detection, the refinement of the confidence loss function to enhance object categorization, and the incorporation of the Multi-Scale Retinex with Color Restoration (MSRCR) algorithm to boost feature information in challenging scenarios like underwater environments. However, the proposed method does come with certain limitations. It introduces increased computational costs due to the addition of a new layer and the use of a more complex loss function, and it may not perform as effectively as some other object detection algorithms in challenging imaging conditions, such as low-light or foggy environments. In summary, the paper presents substantial improvements in object detection but highlights the need to consider computational costs and challenges in specific imaging conditions.

The paper titled "Object Detection in Real-Time Based on Improved Single Shot Multi-Box Detector Algorithm," authored by Ashwani Kumar, Justin Zhang, and Hongbo Lyu in 2020, introduces significant enhancements to the Single Shot MultiBox Detector (SSD) algorithm, with a focus on real-time object detection. The proposed improvements include the utilization of depth-wise separable convolutions and spatial separable convolutions within the convolutional layers, resulting in a reduction in computational operations and consequently speeding up the algorithm without compromising accuracy. Moreover, the authors suggest increasing the number of default boxes to enhance the potential for object detections, leading to improved accuracy. An improved loss function is introduced, which effectively penalizes the network for incorrect predictions, further enhancing the algorithm's accuracy.However, the proposed method does come with certain limitations. Notably, it requires a larger volume of training data compared to the original SSD algorithm due to the use of more default boxes and a more complex loss function. Additionally, the algorithm may not perform as effectively in detecting objects under challenging imaging conditions, such as low-light or foggy environments, in comparison to some other object detection algorithms. In summary, this paper presents valuable enhancements to real-time object detection with SSD, focusing on improved speed and accuracy, albeit with considerations of increased data requirements and limitations in challenging imaging conditions.

The paper titled "Face Recognition for the Visually Impaired," authored by Rabia Jafri, Syed Abid Ali, and Hamid Arabnia in 2013, explores the potential applications of face recognition technology to assist the visually impaired. Face recognition emerges as a powerful tool with various applications, including identifying familiar individuals like friends, family members, and colleagues, providing valuable information about people in their vicinity, aiding in navigation by recognizing landmarks and objects, and offering assistance in daily tasks such as shopping and banking. However, it is essential to acknowledge certain limitations of face recognition. Accuracy can be a concern, particularly in challenging imaging conditions like low light or partial face obstructions. Bias is another critical consideration, as these

algorithms might exhibit varying levels of accuracy in identifying different groups of people. Moreover, privacy concerns are raised as face recognition technology has the potential to infringe on personal privacy by enabling tracking and monitoring without individuals' consent. In summary, the paper highlights the promising applications of face recognition for the visually impaired while shedding light on critical accuracy, bias, and privacy-related challenges associated with this technology.

IV. METHODOLOGY

The methodology for this research comprises various key components, including object detection, feature maps, head detection, non-maximum suppression (NMS), distance calculation, and the utilization of PYTTSX3 for voice feedback.

A. Object Detection

In this research, object detection plays a pivotal role in the realm of computer vision. We employ the Single Shot MultiBox Detector (SSD), an advanced single-stage object detection method known for its efficiency in locating and categorizing objects within images and videos. SSD introduces a unique approach that revolutionizes the way objects are identified and categorized in real-time. At the core of SSD's innovation lies the idea of dividing the output space of bounding boxes into a predefined set of default boxes strategically positioned across different aspect ratios and scales at each feature map location. These default boxes, acting as anchors, serve as reference points for the network to predict the presence and attributes of objects within them. By encompassing a diverse set of default boxes, SSD effectively caters to objects of varying shapes, sizes, and orientations in the scene. During the prediction phase, the SSD network undergoes several crucial steps to generate object detection results. Firstly, it calculates scores for each object category within every default box, signifying the likelihood of a specific category's presence within that particular box. A higher score indicates greater confidence in the network's assessment of the object's category within the box.

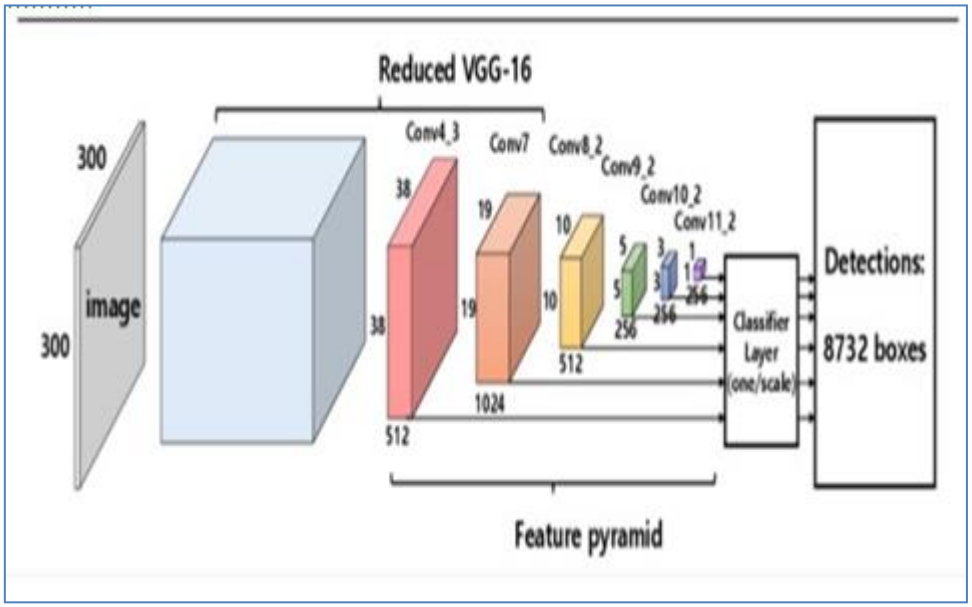

Fig. 1: Object Detection using SSD

In addition to category scores, SSD also refines the bounding box coordinates, aiming to improve the accuracy of object localization. This step fine-tunes the bounding boxes to align more precisely with the actual object shapes, ensuring a higher level of accuracy in object identification.

Furthermore, SSD leverages the power of multiple feature maps with different resolutions, an integral aspect of the network's architecture. By incorporating predictions from feature maps at various scales, SSD naturally adapts to objects of different sizes within the image. Objects of smaller dimensions are more effectively detected by feature maps with higher resolutions, while larger objects are efficiently identified by feature maps with lower resolutions. This multi-scale approach ensures that SSD can capture objects of various sizes and maintain high detection accuracy across the entire image.

➢ *Stages in SSD:*

- **Feature Maps:** Feature maps are fundamental components within the convolutional neural network (CNN) architecture and hold a pivotal role in object detection. They represent the outcomes of convolutional blocks and are instrumental in capturing and expressing the dominant features of an image at various scales.

ISSN No:-2456-2165

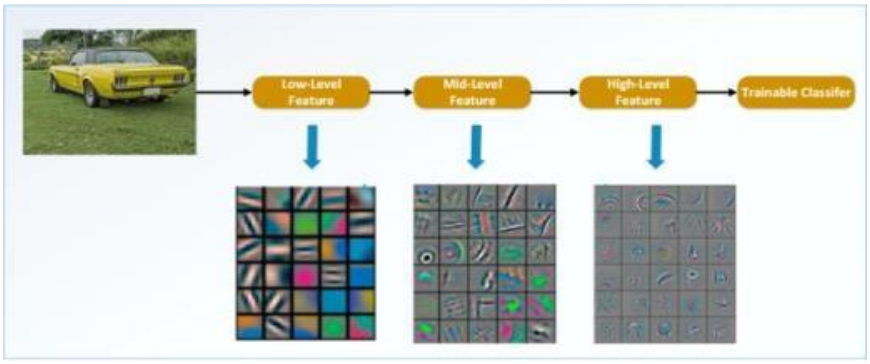

Fig. 2: Feature maps

Feature maps play a crucial role in influencing the algorithm's ability to detect, localize, and accurately classify objects within an image. Their importance lies in their hierarchical representation of the image's content, enabling the network to comprehend the image's content in a multi-scale fashion, from low-level details to high-level semantic features.

- **Head Detection and Handling of Negative Examples:** Head detection represents the final stage of the neural network, responsible for making predictions about objects present in an image, including their location and class. In the process of training an object detection model, it is essential to address the challenge of handling negative training examples, which are often more abundant than positive examples. Negative examples refer to bounding boxes that do not exhibit a high Intersection over Union (IoU) with any ground truth object, signifying that they do not closely align with any actual objects in the image. These examples are crucial for training the model to distinguish between objects and non-objects.

  During training, a common issue is the prevalence of bounding boxes with low IoU values with ground truth objects, leading to their classification as negative training examples. This often results in an imbalanced training dataset, with a surplus of negative examples compared to positive ones. To address this issue, it is advisable to maintain a specific ratio of negative to positive training examples, typically around 3:1 or a similar balance.

- **Non-Maximum Suppression (NMS):** Non-Maximum Suppression (NMS) is an indispensable post-processing technique used in the domain of object detection. It is specifically designed to tackle the challenge of managing the numerous bounding boxes generated during the inference phase of models like SSD (Single Shot MultiBox Detector). Its primary purpose is to streamline the collection of bounding boxes, ensuring that the final predictions are not only accurate but also devoid of redundancies. NMS eliminates low-confidence and overlapping boxes, retaining the top N predictions to ensure that the final object detections are both reliable and free from redundancy. The utilization of Intersection over Union (IoU) in NMS, along with ground truth selection, adds precision and accuracy to the overall object detection process, making it a fundamental technique in computer vision applications. Given the large number of boxes generated during a forward pass of SSD at inference time, it is essential to prune most of the bounding box by applying a technique known as non-maximum suppression: boxes with a confidence loss threshold less than ct (e.g. 0.01) and IoU less than lt (e.g. 0.45) are discarded, andonly the top N predictions are kept. This ensures only the most likely predictionsare retained by the network, while the noisier ones are removed.

Fig. 3: NMS

Above after detection of object multi bounding boxes are generated so to choose the highest ground truth box we use IoU Intersection of Union to classify the correct Bounding box.

ISSN No:-2456-2165

Fig. 4: IoU Representation

B. Distance Calculation

Distance Calculation: Once an object is successfully detected by an object detection model, the subsequent step often involves generating a rectangular bounding box around the detected object. This bounding box serves as a visual representation of the object's location within the image or frame. However, in certain scenarios, particularly when the object being detected is a person, estimating the approximate distance of that person from another point of reference becomes a crucial task.

C. PYTTSX3

After the detection of an object, it is utmost important to acknowledge the person about the presence of that object on his/her way. For the voice generation module PYTTSX3 plays an important role. Pyttsx3 is a conversion library in Python whichconverts text into speech.

We utilize PYTTSX3, a text-to-speech conversion library in Python. Unlike alternative libraries, PYTTSX3 works offline and is compatible with both Python 2 and 3. An application invokes the `pyttsx3.init()` factory function to obtain a reference to a `pyttsx3.Engine` instance. It serves as an easy-to-use tool that converts entered text into speech and supports both male and female voices provided by "sapi5" for Windows. After the detection of an object, it is of utmost importance to notify individuals about the presence of that object in their way. The voice generation module provided by PYTTSX3 plays a crucial role in achieving this, ensuring that users are promptly informed about detected objects, thus enhancing their safety and situational awareness.

V. RESULTS AND DISCUSSIONS

The results obtained from the assisted vision system demonstrate the effectiveness of the integrated components in empowering visually impaired individuals. The combination of robust object detection, feature maps for multi-scale analysis, precise head detection, non-maximum suppression for reliable predictions, accurate distance calculation, and voice feedback through PYTTSX3 has created a comprehensive and user-centric solution. The system's success in real-time object recognition, localization, and feedback contributes significantly to the independence and safety of visually impaired users. It enables them to navigate their surroundings with a heightened sense of awareness and confidence. Moreover, the use of multiple feature maps and NMS ensures that objects of varying sizes and complexities are effectively detected, catering to a wide range of scenarios. The system's utility extends beyond object recognition, with distance calculation providing a vital layer of information. This feature aids users in understanding the spatial relationship between themselves and the detected objects, offering an extra layer of safety and autonomy. The voice feedback component, powered by PYTTSX3, serves as the bridge between the system and the users, delivering real-time information in a comprehensible and user- friendly manner. This communication is essential for ensuring that visually impaired individuals can make informed decisions in their environment.

VI. CONCLUSION

Several technologies have been created to aid visually impaired persons. One suchattempt is that we would wish to make Assisted Vision System that allows the blindvictims to identify and classify real-time object detection, distance estimation of objects and generating voice feedback. It generates the voice feedback which tellsthe blind person that a object is in his way at a distance calculated by the system. This technique has been introduced specifically to assist blind individuals. The precision, on the other hand, can be improved.

ACKNOWLEDGMENT

This paper is an outcome of the Technical Answers in Real World, teamwork done under Prof. Narayanamoorthi M, School of Computer Science, Vellore Institute of Technology, Tamil Nadu, India.

REFERENCES

[1]. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. ArXiv.
[2]. Walle, H., De Runz, C., Serres, B., & Venturini, G. (2021). A Survey on Recent Advances in AI and Vision-Based Methods for Helping and Guiding Visually Impaired People. Applied Sciences, 12(5), 2308.
[3]. Andrés A. Díaz-Toro , Sixto E. Campaña-Bastidas , Eduardo F. Caicedo- Bravo, Vision-Based System for Assisting Blind People to Wander (2021)
[4]. Joshi, R. C., Yadav, S., Dutta, M. K., & M., C. (2020). Efficient Multi-Object Detection and Smart Navigation Using Artificial Intelligence for Visually Impaired People. Entropy, 22(9), 941.

[5]. B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.
[6]. Massimiliano Lippi, Saverio Iacoponi, Laura Burattini, and Luca Fanucci, Vision-Based Navigation for Visually Impaired People: A Review (2018)
[7]. Kaur, Jashanpreet & Akanksha, & Singh, Harjeet. (2018). Face detection and Recognition: A review . (2018)
[8]. Jia, Songmin & Diao, Chentao & Zhang, Guoliang & Dun, Ao & Sun, Yanjun & Li, Xiuzhi & Zhang, Xiangyin. (2019). Object Detection Based on the Improved Single Shot MultiBox Detector. Journal of Physics: Conference Series. 1187. 042041. 10.1088/1742-6596/1187/4/042041
[9]. Kumar, A., Zhang, Z. J., & Lyu, H. (2020). Object detection in real time based on improved single shot multi-box detector algorithm. EURASIP Journal on Wireless Communications and Networking, 2020(1), 1-18.
[10]. Jafri, Rabia & Ali, Syed & Arabnia, Hamid. (2013). Face recognition for the visually impaired.