IMPLEMENTATION 35OF A VOICE-GUIDED WEARABLE DEVICE FOR LOCATING THE PERSONAL BELONGINGS OF VISUALLY IMPAIRED PEOPLE

SWATI SHILASKAR[1] , SHRIPAD BHATLAWANDE[2] , SHRRUTI SURANJE[3]

[1]Faculty, Dept. of E&TC Vishwakarma Institute of Technology Pune, India
[2]Faculty, Dept. of E&TC Vishwakarma Institute of Technology Pune, India
[3]Student, Dept. of E&TC Vishwakarma Institute of Technology Pune, India

E-mail: [1]swati.shilaskar@vit.edu, [2]shripad.bhatlawande@vit.edu, [3]shrruti.suranje20@vit.edu

ABSTRACT

The Voice-Guided Indoor Assistance System is an innovative solution to help visually impaired individuals navigate indoor spaces. It uses Bluetooth communication and image processing techniques to locate objects and provide precise navigation instructions. The system also has a user-centric design and a unique capability to locate the device through an assistive cane. With an impressive 98% accuracy, the system ensures real-time responsiveness in diverse lighting conditions. By empowering individuals with visual impairments to navigate indoors confidently and independently, this technology marks a significant leap forward. It enhances autonomy and engagement in real-world scenarios, and the seamless integration of this system is a pivotal stride towards creating an inclusive environment for everyone.

Keywords: *Visually Impaired, Indoor Navigation, Object Localization, Computer Vision*

1. INTRODUCTION

Among the five sensory organs that enable us to perceive and become familiar with our surroundings, sight accounts for about 80% of our perceptions of the world. It is important for keeping us safe even when other senses such as smell and taste are not working. According to a survey conducted by WHO, about 135 million people worldwide have poor eyesight, and approximately 40 to 45 million people are blind. This has a significant impact on their daily lives. Visually impaired people may have difficulty performing routine tasks such as finding personal belongings.

To address this issue, several machine learning-based systems have been developed previously. Hassan et al. [1] presented an AI-based system to assist the visually impaired with navigation that detects and classifies objects in real time using OpenCV and deep learning algorithms. The YOLOv3 object detection system is used in this system, which provides audio feedback to the user using GTTS. Nasreen et al. [2] developed a YOLO system that consists of a website that functions as an interface for capturing images from a smartphone camera. These captured images are sent to a web server where the ML model is used to detect objects in the image. The results are then communicated back to the client, where a Google Voice converts the text into speech and narrates the object details to the user. Similarly, the systems in [3] also utilize the YOLO algorithm and the COCO dataset to train the YOLO model for object detection, OCR to recognize textual data placed before the camera module, and text-to-speech conversion using GTTS or eSpeak and pyttsx3.

This paper's proposed model advances beyond previous research by integrating object detection with navigation. This creates a more comprehensive indoor navigation system for the blind. The system leverages machine learning to recognize objects within indoor environments and guides visually impaired users through audio feedback.

The proposed system consists of two main components: feature extraction and object recognition. Feature extraction is performed using ORB (Oriented FAST and rotated BRIEF), while classification is performed using Decision Trees (DT), Random Forests (RF), and K-Nearest Neighbours (KNN) classifiers. We will explore machine learning-based object detection techniques to identify indoor objects and provide audio feedback to the user. The technique also investigates the feasibility of integrating a navigation function with an object detection system to help the user reach a desired destination. The system improves the living conditions of visually impaired individuals by providing them with an efficient and reliable indoor navigation system through voice assistance.

The paper is divided into the subsequent parts, part II gives an outline of the work related to object detection and navigation for the blind, part III describes the proposed system methodology, followed by experimental results in part IV, and concludes work in the V part.

2. LITERATURE SURVEY

This literature survey delves into recent research works that have explored object recognition systems with a focus on key components such as object detection, audio feedback, and remote navigation. These systems merge advanced technologies—computer vision, and wearables—to completely address the unique hurdles faced by individuals with visual impairments.

One proposed method for object recognition is described in [4], the authors develop a system which utilizes TensorFlow object detection API and SSDLite MobileNetV2 model trained on COCO dataset for object detection. Gradient particle swarm optimization is employed to optimize model layers and hyperparameters. Real-time object detection is performed using a Raspberry Pi 4B microcontroller and Pi camera integrated into a head cap for efficient obstacle detection. Additionally, a secondary model, "ambiance mode," is trained on a weather dataset to describe surroundings in detail. Chaccour et al. [5] present a solution for indoor object detection and Human Machine Interface (HMI) navigation through a smartphone app connected to indoor WLAN. They compare video files with indoor ceiling IP camera images through computer vision. In [6], authors offer a technique to aid visually impaired individuals in navigating indoor spaces using a graph-based internal map, Dijkstra's algorithm for optimal path calculation, and continuous navigation commands. The system employs camera calibration for accurate localization and demonstrates that ArUco markers are better suited than QR codes.

An upgraded YOLOv5s model [7] for better object detection in blind navigation systems. Integrating a feature scale and prediction head, it enhances small object detection on blind paths. Incorporating a spatial knowledge distillation (SKD) technique in the feature fusion stage: This technique facilitates the model to selectively focus on informative features from different scales, resulting in more precise object extraction irrespective of their distance and size. Experimental results show a 6.29% mean average precision (mAP) improvement over YOLOv5, especially excelling in detecting smaller objects like cats, dogs, and fire hydrants.

For instance, the system in [8] utilizes Pi-Camera-equipped glasses and a neural network to detect objects and provide audio feedback via the Espeak text-to-speech engine. Trained on 200 images of 10 household objects, the system achieves a precision of 0.85 and a recall of 0.8. The device in [9] employs artificial intelligence and image processing to detect faces, colors, and objects, alerting the user through sound or vibration. OpenCV and Python are used for programming and implementation, entailing dataset creation, Haar Cascade Algorithm-based training, and face recognition with another program. Another proposed navigation system incorporates a Raspberry Pi board, heartbeat, ultrasonic sensors, camera, push button, GPS, and GSM modules [10]. The camera captures the visually impaired user's surroundings, the ultrasonic sensor detects obstacles, and the heartbeat sensor monitors vital signs. Object recognition employs a frozen inference model and a Text-to-Speech Synthesizer.

Advancements in location and navigation technologies are revolutionizing the way we perceive and interact with indoor spaces, particularly benefiting groups like the visually impaired. The study [11] presents a navigation system designed to assist individuals with severe visual impairment in indoor building navigation. Leveraging a camera installed on either a mobile phone or a wearable device, the system employs visual cues like doors, stairs, signage, and fire extinguishers as landmarks to aid users in navigating their surroundings. This real-time system detects obstacles and dynamically adjusts the route. Initial testing of the system yielded positive outcomes, and ongoing efforts aim to enhance its precision while expanding the repository of recognized landmarks. Adaptive aids for blind individuals that utilize CV and deep learning techniques have been proposed in several studies. Authors in [12] comprehensively examines indoor positioning systems (IPSs) with a specific focus on their role in assisting visually impaired people with indoor navigation. It categorizes IPSs into three main architectures: self-positioning, infrastructure positioning, and assisted by self-directed infrastructure. The review emphasizes the challenges posed by signal loss within indoor environments and proposes various technologies and methodologies to enhance both accuracy and scalability. The study highlights the critical aspects of adaptability, accommodating diverse hardware, and power consumption in IPS design. Utilizing Harris Corner Detection (HCD) from Computer Vision in [13] to predict object positions and distances by analyzing image corners. Achieving 88% accuracy in object detection, the smartphone-based solution captures, processes, and converts images into audio feedback, facilitating real-time navigation for visually impaired individuals. Incorporating the Triangle Rule for distance estimation, it shows promising results in practical testing.

A navigation system detailed in [14] conducted a thorough review of indoor navigation technologies. They examined non-vision-based methods using technologies like RFID, infrared, and UWB, as well as vision-based solutions including single-camera systems (QR code, ARUCO) and 3D cameras (RGB-D, ToF). The study highlighted the limitations in existing research, such as limited inclusion of visually impaired perspectives, small datasets, absence of standardized evaluation methods, and limited use of advanced algorithms. Barontini et al. [15] presented a user-centered approach to indoor navigation using wearable haptics and obstacle avoidance algorithms. They used a haptic feedback device on the user's wrist to provide tactile guidance and integrated obstacle avoidance algorithms. The system achieved an accuracy of 90%, but challenges remain in differentiating moving and fixed obstacles and addressing hardware architecture and lighting conditions. Kandalan and Namuduri [16] discussed indoor navigation system modules, including wayfinding, obstacle avoidance, and human-machine interaction. They explored techniques like GPS for outdoor navigation and trilateration for indoor positioning using WLAN access points. The paper emphasized the need for accurate sensors, adaptable algorithms, and user-friendly systems.

Article [17] focuses on prototype research rather than commercial solutions. It highlights challenges with current assistive devices, emphasizing the acceptance of smartphone-based interfaces among VI users. Ultrasonic sensors on white canes or attire are effective for detecting obstacles within 5-20 meters. Non-camera-based technologies like UWB, BLE, and RFID offer varying levels of accuracy and cost-effectiveness. Factors like accuracy, cost, and ease of deployment are all crucial. The authors conclude that UWB offers the most accurate positioning but may be cost-prohibitive for widespread use. BLE presents a more affordable option but with slightly less accuracy. Hybrid systems combining multiple technologies offer improved accuracy but are costlier. The paper [18] introduces VIPEye, a system aiding visually impaired people in navigating new environments independently. It uses graph mining and computer vision to create a Safe & Interesting Path (SIP) from start to finish, guiding users via voice commands. It discusses VIPEye's design, challenges in development, like task complexity, and suggests future research directions for improving mobility services for the visually impaired. Moreover, another paper [19] proposes a new comprehensive dataset designed to evaluate the performance of object, face, and proximity detection systems used in assistive technologies for visually impaired people. The dataset incorporates various real-world scenarios with diverse lighting conditions and complexities to mimic everyday situations. It includes annotated images with corresponding ground truth data to facilitate training and assessment of machine learning models for accurate object and face recognition, as well as for estimating object proximity. This dataset aims to improve the development of assistive technologies by providing a more realistic and comprehensive evaluation benchmark. The study [20] proposes a method to estimate 3D location of chair objects on the floor surface using a single image and a perspective grid approach. The system achieves an average accuracy of 6.47 centimeters which is significant compared to the object dimension. The biggest factors affecting accuracy are lighting condition and constant parameter determination. The authors plan to improve the system by allowing for multiple object categories, applying the method on videos, and developing category-specific object datasets.

Extensive research has been directed toward indoor navigation and obstacle detection utilizing an array of sensors. However, a significant research void exists concerning the facilitation of visually impaired individuals in indicative indispensable personal items within indoor environments. Vision-based localization technologies require line-of-sight (LOS) and are sensitive to environmental conditions, limiting their effectiveness in complex indoor environments. Some technologies, such as ultrasound-based systems and ultra-wideband (UWB) signals, offer high accuracy but are associated with high costs, making widespread deployment challenging. Installation and maintenance costs of positioning infrastructure, such as RFID readers and beacons, may pose barriers to the adoption of indoor navigation systems, especially in large-scale environments. The insufficient focus on personal belongings recognition in existing assistive systems for visually impaired individuals highlights a significant gap in addressing the specific needs and challenges faced by this user group. While many systems utilize general object detection techniques, they often overlook the importance of identifying and recognizing items that are essential for daily activities and mobility, such as chargers, mobile phones, electric sockets, and canes.

Recognizing personal belongings is crucial for visually impaired individuals to maintain independence and navigate their surroundings effectively. Unlike generic objects, which may vary widely in shape, size, and appearance, personal belongings tend to have consistent and recognizable features that are familiar to the user. Therefore, developing object detection models tailored to identify these specific items can greatly enhance the usability and effectiveness of assistive systems for visually impaired individuals. This paper seeks to propose an innovative computer vision-driven framework specifically designed to aid visually impaired individuals in the efficient identification and retrieval of common items, thereby contributing to an improved quality of life. Moreover, incorporating personal belongings recognition into assistive systems can improve user confidence and reduce reliance on external assistance. Instead of having to rely on sighted individuals or tactile exploration to locate specific items, visually impaired individuals can navigate their way to objects like mobile phones, chargers, canes, or sockets which are necessary for their daily routines.

3. METHODOLOGY

The system was designed to assist visually impaired individuals. This system was designed to capture and process images using voice commands, perform object detection and localization, and provide users with voice-guided directions to their intended objects. The proposed object identifier was developed to recognize the user's personal belongings, including items like a charger, mobile phone, electric socket, and cane. An innovative aspect of the system was its ability to locate a cane or the device itself through a Bluetooth-activated buzzer. This multifaceted approach improved indoor navigation and object retrieval, enhancing the autonomy and mobility of visually impaired individuals.

The architecture aimed to empower visually impaired individuals in their daily lives, integrating camera input, voice commands, Bluetooth connectivity, and auditory feedback for a user-centric experience. The approach included two main sections: A) Computer vision-based object identification and locator. B)Bluetooth-based cane locator.

3.1 Computer Vision-Based Object Identification And Locator

Computer vision-based component of the model emphasizes on the system's proficiency in identifying and locating objects through advanced image processing techniques. The hardware implementation of the suggested object locator model incorporates a standalone device using Raspberry Pi 4B (RPI 4B), chosen for its robust performance and versatility. A systematic view of the implemented device’s hardware circuit is presented in figure 1. A battery source is connected to ensure the system's independence and portability, achieved through a power bank. User interaction is further streamlined with the inclusion of buttons on the RPI. These buttons serve specific functions, with button 1 acting as the device's start button for activation, and button 2 allowing the user to turn off the buzzer when actively searching for the device with the assistance of the cane, as illustrated in figure 7.

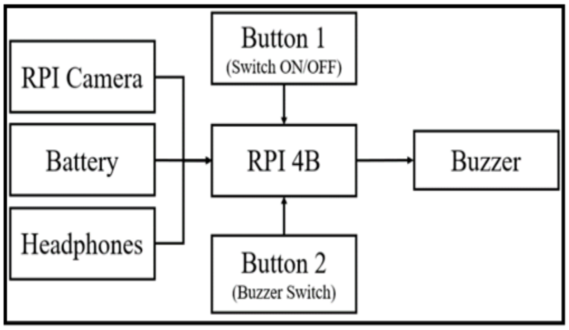

Figure 1: Schematic Representation Of System Hardware Architecture

The complete system block diagram, shown in figure 2, illustrates the operational flow. This flow encompasses image capture and processing in response to voice commands, and object detection specifically for mobile chargers, mobile phones, electric sockets, and the cane. Additionally, the system achieves object localization and guides users precisely through voice prompts to their intended objects.

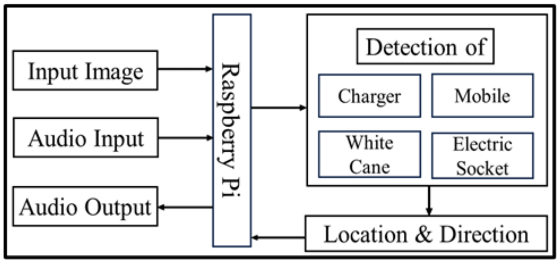

Figure 2: Overview Of Implemented Object Locator System

The system's implementation seamlessly integrates multiple critical components. The system captured and processed images through voice commands, conducted object detection and localization, and guided users to their desired objects using voice prompts. Upon initiation, the system interacts with the user using voice commands like “Which object are you looking for?”, articulating their desired object. The system promptly captures an image using the device's integrated camera. This captured image is then transmitted to the Raspberry Pi (RPI) for further processing. Within the RPI, the image undergoes a series of crucial processing stages. Initially, a preprocessing step optimizes the image's quality, preparing it for subsequent analysis. Subsequently, the system employs the ORB feature extraction technique, chosen for its significant excellence. The important phase lies in clustering, utilizing the K-means algorithm with k=6 to effectively group akin features. This clustering proves instrumental in facilitating the recognition of objects, and each cluster is duly associated with predefined object labels.

The core of the system's operation lies in its classification capabilities. When the system successfully identifies the desired item, it generates an audio signal as a user notification. To enable this functionality, we maintain a dataset containing images of the user's items, serving as a reference for the system. This dataset encompasses multiple images of each item, captured under various conditions, including different camera angles, scales, lighting scenarios, and potential obstructions. Here, the system identifies the object the user requests and determines whether it exists within the dataset. If the desired object is not found, the system provides a vocal prompt, articulating the unavailability of the requested item. In contrast, if the object is identified within the dataset, the system proceeds to the object localization phase. Object localization serves to validate the presence of the requested object within the captured image. In instances where the object eludes detection, the system issues a voice command, courteously notifying the user of its absence. Conversely, if the object is successfully identified, a process of non-maximum suppression (NMS) refines the object's bounding box. The localization process is equipped to handle scenarios involving multi-labelled images. When the user requests the detection of a specific object among multiple objects in the scene, the system employs a prioritization mechanism to locate the object of the highest priority as per the user's request.

The system then adeptly issues voice-guided navigation instructions that are text-to-speech conversion to the located object. The system takes advantage of the centroid of the refined bounding box. By determining the centroid's position relative to the user, the system provides clear and concise voice-guided navigation instructions. For instance, if the centroid is towards the left, the user is directed to turn left. Similarly, if the centroid is to the right, the user is instructed to turn right. In cases where the centroid is centrally located, the user receives guidance to move forward. This intuitive approach enhances the user's mobility and engagement with the environment. Following a brief pause of 10 seconds for the user's convenience, the system re-engages by soliciting a fresh object specification. This iterative and user-centric approach ensures an efficient and dependable assistive experience for individuals with visual impairments.

3.1.1 Visual input and voice command

The core of the system's functionality relied on capturing the user's surroundings in real-time. This was achieved through a Raspberry Pi camera, connected via the CSI camera port. The camera was a source of input images, allowing the system to perceive the environment visually. To utilize the system, users issue voice commands to request the location of a specific item. Subsequently, they wear the camera-based object locator system to search for the requested item. To enable intuitive and accessible interaction, the system was equipped with voice command capabilities. Visually impaired users could issue voice commands to request the location of specific items, initiating the object identification and localization process. These voice commands were adeptly captured by headphones connected to the RPI through a 3.5mm audio jack. This seamless integration of voice input ensured that users could seamlessly communicate their requests to the system.

3.1.2 Object Detection and Classification

It comprises several key components, commencing with the model training process for object detection and classification, which includes Dataset Acquisition and Preprocessing, Feature Extraction, Feature Clustering, and Feature Classification. The complete flow of the implementation is visualized in figure 3.

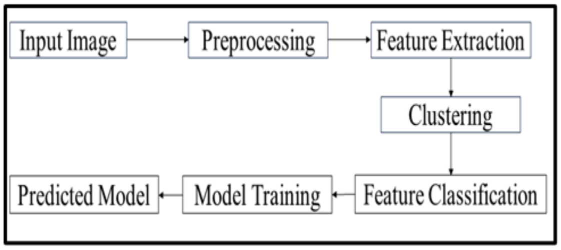

Figure 3: Model Training Process Diagram For Object Identification

3.1.2.1 Data acquisition and preprocessing

The training dataset for the proposed model was precisely curated. The dataset comprises images from four categories: mobile charger, mobile, cane, and electric socket. This dataset encompasses multiple images of each item, captured under various conditions, including different camera angles, scales, lighting scenarios, and potential obstructions. Sample images from the dataset, representing various categories, are illustrated in figure 4.

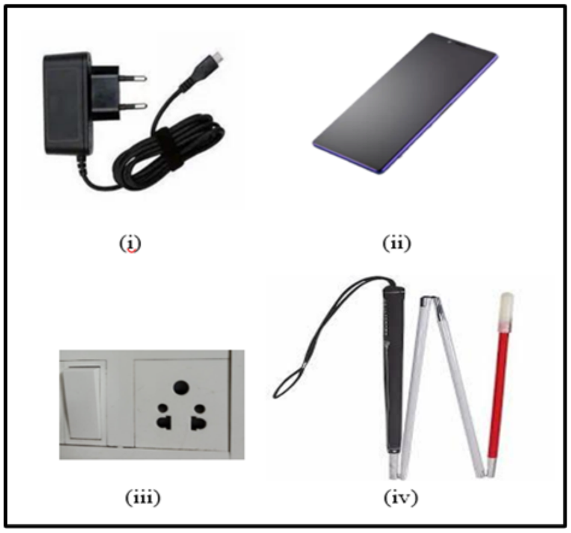

Figure 4: Illustrates Sample Images From The Dataset, Representing Various Categories (i) Mobile Charger, (ii) Mobile, (iii) Electric Socket, (iv) Cane.

In total, 8,000 photographs were captured, with each class containing 2,000 images as shown in table 1. The images captured by the two cell phones had varying resolutions, therefore, to ensure uniformity and reduce computational complexity, all images were resized to 250 x 250 pixels.

Table 1: Details Of Dataset

Class	Number of Images
Mobile Charger	2,000
Mobile	2,000
Electric Socket	2,000
Cane	2,000

The preprocessing pipeline began by converting the RGB images into grayscale, aligning with established practices for subsequent analyses. To enhance feature extraction, Prewitt edge detection was systematically applied to these grayscale images. The application of Prewitt edge detection aimed to highlight important boundaries and contours, facilitating subsequent feature extraction processes.

3.1.2.2 Feature extraction

It is a crucial process that involves the identification and quantification of distinctive patterns, structures, or attributes within images. This procedure serves to encapsulate essential visual information while simultaneously mitigating data complexity. Our model, the Oriented FAST and Rotated BRIEF (ORB) feature extraction technique was evaluated for its efficiency. ORB, a variant rooted in the widely recognized Scale-Invariant Feature Transform (SIFT), was initially assessed for its potential to augment the model's performance.

Operating on the principle of identifying key points in images, ORB employs Binary Robust Independent Elementary Features (BRIEF) descriptors to articulate these points. These descriptors capture indispensable image information, endowing the model with robustness against diverse transformations, including rotation and scaling. ORB consistently achieved high accuracy rates when used in conjunction with different classifiers. Key points are identified based on intensity differences, ensuring that only salient features contribute to the subsequent steps of the algorithm. These key points play a crucial role in forming the basis of the feature descriptors. Equation (1) calculates the intensity difference (d_ij) between two-pixel locations (x_i ,y_i) and (x_j ,y_j). In the framework of ORB, it is utilized in key point detection and description. A significant intensity difference is indicative of a distinctive point in the image.

$$d_{ij} = I(x_i, y_i) - I(x_j, y_j) \quad (1)$$

The FAST (Features from Accelerated Segment Test) algorithm, described by (2), is employed for key point detection. It defines the criteria for considering a pixel as a key point based on the intensity relationships with its surrounding pixels. This equation is used in the initial phase of the ORB algorithm to identify locations in the image that are distinctive and possess unique features. These key points become essential for subsequent orientation assignments and descriptor generation. A pixel p is considered a key point if there exist n contiguous pixels in a circle of 16 surrounding pixels that are either all brighter or all darker than the intensity of p.

$$\mathrm{k} = \sum_{k=1}^{16} \sin\left(I_p - I_{p_k}\right) = n \quad (2)$$

Orientation assignment is a key aspect of refining the features detected in the previous steps. Equation (3) assigns an orientation (θ) to each key point based on the intensity centroid of surrounding pixels. This step is crucial for ensuring the rotational invariance of the features. The assigned orientations contribute to the creation of descriptors that are invariant to image rotations.

$$\theta = tan^{-1}\left(\frac{\sum_p w_p\,(y_p - \bar{y})}{\sum_p w_p\,(x_p - \bar{x})}\right) \quad (3)$$

Where w_p represents a weight assigned to each pixel, and (x ,y) denotes the centroid.

These equations, represented as (1), (2), and (3), encapsulate the fundamental steps in the ORB algorithm, providing a clear understanding of its feature extraction process tailored for the object identification dataset.

3.1.2.3 Clustering

To effectively group similar features, we employed the K-Means clustering algorithm, which partitions data into distinct clusters, each characterized by its centroid. The choice of the number of clusters, determined through analysis like the elbow method, was substantiated with a value of 6. After fitting the entire feature dataset into the K-Means model, labels were assigned to each image based on their proximity to cluster centroids. Each cluster formed during the analysis was associated with specific labels capturing object characteristics. Objects with similar features naturally gravitated towards the same cluster, establishing a direct link between feature similarity and object identity. In large datasets, clustering serves as a valuable data reduction technique, grouping data into clusters for manageability and reduced computational complexity. Our study conducted a comparative analysis of two clustering algorithms: K-Means and Gaussian Mixture Models (GMM). The elbow method, illustrated in figure 5, determined the optimal number of clusters for each algorithm.

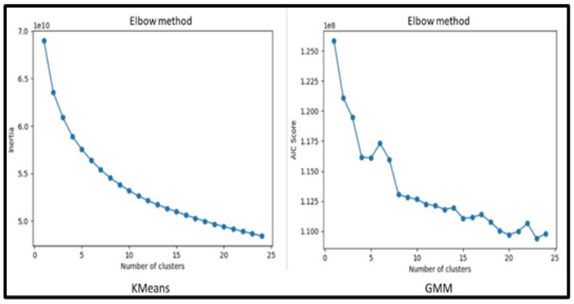

Figure 5: Elbow Plot Using Kmeans And GMM

For GMM, the Akaike Information Criterion (AIC) was utilized. At the same time, the Within-Cluster Sum of Squares (WCSS) was employed for K-Means.

3.1.2.4 Classification

Moving beyond clustering, our research integrated a robust classification phase crucial for recognizing and distinguishing objects. While clustering organizes data based on similarities, it lacks the inherent capability for object recognition or labelling. In contrast, classifiers like Decision Trees (DT), Random Forests (RF), and K-Nearest Neighbors (KNN) are explicitly engineered for these tasks, learning from labelled data to make predictions on new, unlabeled data points. This classification process utilized clustered data to enhance the classifiers' ability to differentiate between objects based on unique patterns and features, significantly improving object recognition within our assistive system. DT constructs a tree-like structure by partitioning the input space based on features, while RF combines predictions from multiple trees for enhanced precision and reliability. KNN classifies new instances by considering their k-nearest neighbors, with an initial value of 13 used in this study. Models were trained on extracted features, and their performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.

3.1.3 Object localization

It represents the phase post-clustering and model training, serving as a cornerstone in image processing. Its significance lies in its ability to accurately identify specific objects within images, even in cases where multiple labels exist. This section delves into the methodologies and fundamental contributions of object localization. The primary aim of Object Localization is to precisely identify and locate specific objects within captured images. This objective demands a suite of algorithms and techniques, such as the Image Pyramid, Sliding Window, and Non-Maximum Suppression (NMS). These components are instrumental in achieving exceptional precision and robustness in object identification and navigating the complex landscape of images with multiple labels. Initialization of the object localization involves setting up the sliding window's step size, defining the minimum size prerequisites for generating the image pyramid, and establishing a minimum confidence threshold, crucial for robust object identification.

3.1.3.1 Image pyramid generation

The algorithm initiates by crafting an image pyramid, an organized hierarchical representation of the original image at various scales. This hierarchical structure begins with resizing the initial captured input image and ensuring the resized images comply with predefined size constraints. This multi-scaled representation allows intricate analysis across different granularities, enabling precise detection of objects with varying sizes and complexities within the image. Algorithm 1 aligns with the common practice of using the image pyramid first to create multiple scales of the image, then applying the sliding window technique represented in Algorithm 2 to each patch within the pyramid to detect objects at various scales.

Algorithm 1: Apply Image Pyramid

```
Input: Image
Output: Resized patches in descending scales
        Initialization
        minSize ← image pyramid minimum size
          while: true
            if resized image < minSize
              break
            end
          return next scaled image
          end
```

A patch refers to a smaller subsection or fragment of an image. In the context of image processing or computer vision, when an image is divided into smaller, usually rectangular, sections, each of these sections or subsections is often referred to as a patch.

3.1.3.2 Sliding windows and object localization

It is represented in Algorithm 2. Within each layer of the image pyramid, the algorithm systematically traverses the image using a sliding window approach. Vertically and horizontally iterating through these windows, it systematically extracts coordinates for the top-left corner of each window. Employing object localization techniques, within each window, the algorithm precisely determines the object's coordinates within these windows. It carefully examines the content within these patches to get the exact coordinates where the detected object resides. The goal is to precisely identify the presence and location of the object of interest. It selectively stores coordinates and their probabilities in a structured dictionary, contingent upon surpassing the predefined minimum confidence threshold.

The minimum confidence threshold, typically set at 0.6, plays a pivotal role. It acts as a filter, allowing the algorithm to retain only those object detections that exceed this confidence level. If the detected object's probability, derived through the object localization process, surpasses this threshold, the algorithm stores the coordinates and associated probability in a labels dictionary. The significance of the minimum confidence threshold lies in its ability to control the accuracy of the detections.

Algorithm 2: Apply Sliding Windows and Object Localization

```
Input: Image
Output: Object coordinates in the image
Initialization
step_size ← sliding window
min_confidence ← minimum confidence threshold
for image in Image Pyramid do
    for window in generated sliding windows do
        Iterate vertically and horizontally with step_size
        Compute coordinates (x, y) for window's top-left corner
        Apply Object Localization:
            Input: Window
            Output: (x, y) ← coordinates of detected object

        if probability >= min_confidence then
            Store coordinates and probability in labels dictionary
            labels[label].append((box, probability))
        end
    end
end
```

Increasing the threshold can lead to higher precision but might discard valid detections with slightly lower probabilities. Conversely, reducing the threshold might capture more detections but may include false positives, reducing the overall precision of the detected objects. Adjusting this threshold is crucial in balancing precision and recall, catering to the specific needs of object detection application.

3.1.3.3 Non-maximum suppression (NMS)

Following the object localization task, multiple bounding boxes are generated around potential objects within an image. Each bounding box is associated with a confidence score, typically representing the likelihood that the box contains an object of interest. These bounding boxes are inputs to the NMS process. Central to NMS is the concept of Intersection over Union (IoU). IoU is a mathematical metric that quantifies the overlap between two bounding boxes. It is calculated as the ratio of the area of intersection between two boxes to the area of their union. The equation for IoU is represented (4).

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \qquad (4)$$

NMS requires setting an IoU threshold, defining acceptable overlap between bounding boxes. To prevent duplicate detections, overlapping bounding boxes exceeding a specific degree of overlap (IoU threshold) are eliminated. Algorithm 3 is designed for applying Non-Maximum Suppression (NMS), a technique crucial in refining and filtering detected objects within an image.

Algorithm 3: Apply Non-Maximum Suppression

```
Input: Multiple bounding box coordinates generated from object localization
Output: Final object coordinates in the image
Initialization
step_size ← sliding window
boxes[] ← coordinates of the bounding boxes of detected objects
proba [] ← confidence scores of each bounding box identified in image
nms() ← Non-Maximum Suppression function
for label in labels.keys() do
    boxes ← np.array([p[0] for p in labels[label]])
    proba ← np.array([p[1] for p in labels[label]])
    boxes ← nms(boxes, proba)
end
```

The algorithm operates based on the step_size parameter, controlling the sliding window's dimensions during object detection across the image. Within the iteration through each label in the labels dictionary, '*boxes*' compile the bounding box coordinates for detected objects, structured in a NumPy array. Concurrently, '*proba*' collects the confidence scores associated with each bounding box, organized as a NumPy array. These arrays undergo processing within the '*nms()*' function, the algorithm's core, The algorithm operates based on the step_size parameter, controlling the sliding window's dimensions during object detection across the image. Within the iteration through each label in the labels dictionary, 'boxes' compile the bounding box coordinates for detected objects, structured in a NumPy array. Concurrently, 'proba' collects the confidence scores associated with each bounding box, organized as a NumPy array. These arrays undergo processing within the 'nms()' function, the algorithm's core, executing logic to filter redundant and overlapping detections. Through this step, the algorithm ensures that the object coordinates retained accurately represent unique and non-redundant objects within the image. This systematic process refines detections using NMS, eliminating redundancies and overlaps, resulting in a more accurate representation of distinct elements within the image.

The assessment of each bounding box is accomplished by analyzing its confidence score, which must be equal to or greater than 0.6 for this specific scenario. The algorithm begins by selecting the bounding box with the highest confidence score as the primary detection. It then compares the IoU value of this primary detection with all other bounding boxes. Any bounding box with an IoU value above the threshold is suppressed, while the primary detection is preserved. This process is demonstrated in figure 6 as bounding box refinement.

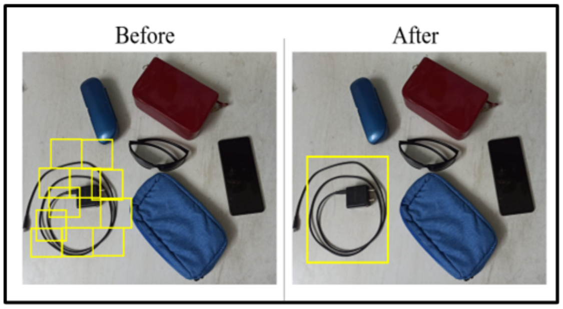

Figure 6: Bounding Box Refinement Using NMS

This iterative procedure ensures that the most confident and non-overlapping bounding boxes are retained. While eliminating overlapping ones, NMS refines object localization results leading to precise and non-redundant object detections in complex scenes. This is essential for fine-tuning and optimizing object detection models, ultimately enhancing their accuracy and reliability.

3.1.4 Voice-guided navigation

To facilitate effective communication and provide crucial feedback to the user, both the RPI and the assistive cane were equipped with 5V Passive Buzzers. These buzzers generate auditory signals, allowing the device to convey information and guide the user effectively. In this section, the voice-guided navigation process, a crucial component of the proposed assistive system, is explored in detail. This process involves the conversion of voice inputs into text, enabling the system to understand and process the user's spoken commands effectively. When initiated, the device actively engages with the user by asking, "Which object are you looking for?" The user's spoken response is then converted to text for further processing. The system, as part of its user-centric design, seamlessly converts text to speech for issuing navigation instructions to the user. This text-to-speech conversion enhances the system's accessibility and ensures users receive clear and concise auditory guidance through headphones attached. This process involves the conversion of text to speech, enabling the system to communicate clear and concise auditory instructions to the user. The choice of this text-to-speech conversion method was driven by its capability to enhance the accessibility of the system, ensuring that users receive navigational guidance in a comprehensible and efficient manner.

At the core of the voice-guided navigation system lies the centroid of the refined bounding box. This centroid serves as a critical reference point for determining the precise location of the requested object within the captured image. The centroid's calculation is based on the geometric properties of the object's bounding box, and it holds substantial importance in providing accurate navigation instructions. Through this process, the system can reliably guide users by indicating the direction in which they should move to reach the desired object. The voice commands issued to users with the utmost consideration for clarity and effectiveness. These voice commands are generated based on the calculated position of the centroid relative to the user's perspective. For instance, if the system detects the centroid on the left side of the user's field of view, it articulates the command, "Turn left," prompting the user to change their orientation accordingly. Similarly, when the centroid is situated on the right, the instruction is straightforward: "Turn right." In cases where the centroid aligns centrally, the system issues the command, "Move forward," facilitating the user's onward movement.

The technical sophistication of this intuitive approach lies in its ability to precisely interpret the object's location through the centroid and convert this information into actionable voice commands. This not only enhances user mobility but also instills a sense of independence and confidence in visually impaired individuals. By bridging the gap between the digital representation of the environment and auditory instructions, this voice-guided navigation system contributes significantly to improving the quality of life for its users.

3.2 Bluetooth-Based Cane Locator

The proposed model introduces an innovative two-way communication mechanism aimed at significantly enhancing the mobility and independence of individuals with visual impairments. This innovative approach focuses on interaction between the user's primary object identifier and locator device and an assistive cane. Both devices incorporate a dedicated button that activates a shared buzzer, enabling localized auditory signals. This distinctive feature enables visually impaired users to locate either device, regardless of which one they hold at any given time. The assistive cane includes a button to locate the device, further enhancing user autonomy and mobility.

Facilitated by widely used Bluetooth technology, the interaction between the user's device and the assistive cane enables seamless two-way communication. The Raspberry Pi 4, a key component of the proposed system, features built-in Bluetooth functionality, simplifying the connection process and emphasizing user convenience and technological efficiency. Utilizing this the object locator device establishes a robust and versatile communication channel. It connects to an HC-05 Bluetooth module integrated into the assistive cane, creating a bidirectional communication link. This integration significantly aids in localizing the device, thereby enhancing overall user independence and mobility. This mechanism operates through the assistive cane's button activation, establishing a Bluetooth connection with the primary object locator device. This prompts the device's buzzer to activate, aiding the user in locating it using the cane figure 7.

The two-way communication system extends to the user's ability to locate the assistive cane using the primary device shown in figure 8, or vice versa further enhancing the user's autonomy and engagement with the assistive system. Visually impaired users can trigger the buzzer on either the device or the cane, allowing them to locate either of the objects. The interaction is swift, intuitive, and non-reliant on complex navigation systems, further accentuating the system's user-friendly design.

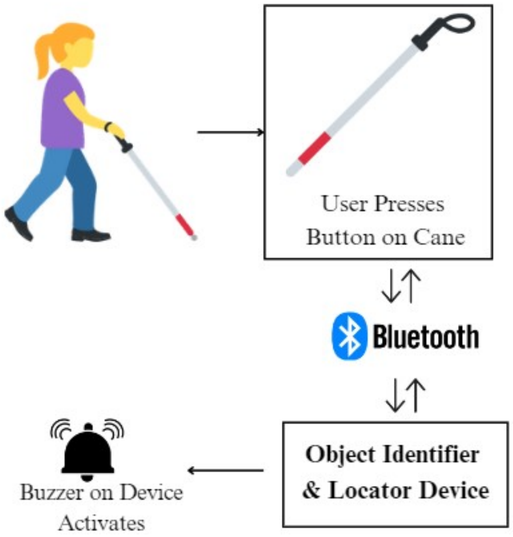

Figure 7: Finding The Object Locator Device Using The Assistive Cane

By providing users with the means to independently locate their device or assistive cane, our research underscores a commitment to actively engage users with their surroundings.

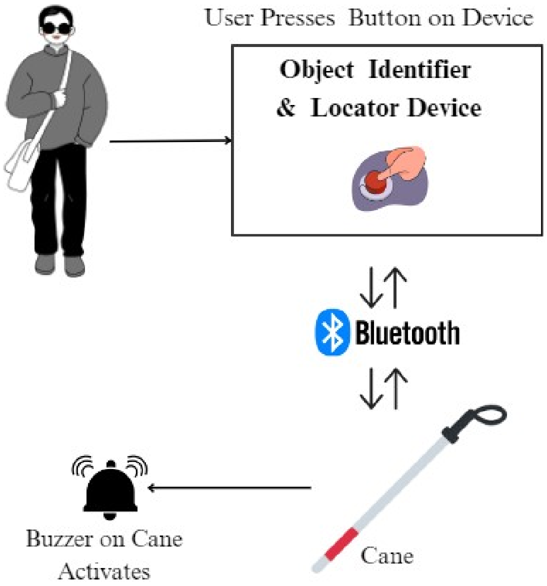

Figure 8: Blind Users Can Find The Assistive Cane Using the Proposed Object Locator Device.

It is a novel contribution, reinforcing the user-centric design of our assistive system while providing a practical and efficient solution to mobility challenges.

4. RESULTS AND DISCUSSION

After The proposed system is a vision-based indoor assistance personalized to assist visually impaired individuals in locating objects of interest within their indoor environment The system's design revolves around the feature extraction method by ORB followed by clustering techniques KMeans and GMM included, complemented by a range of classifiers. For the task of classification, a trio of classifiers DT, RF, and KNN were used. The comparison of the accuracy of all combinations and classifiers is shown in table 2.

Table 2: Results Of ORB Algorithm

Clustering Techniques	Classifiers		
	DT	RF	KNN
KMeans	76.8%	98.4%	98.1%
GMM	65.4%	78.7%	76.6%

The analysis of the tables reveals that the ORB feature extraction method combined with the K-means clustering technique exhibited the highest overall accuracy. Among the three available classifiers, Random Forest emerged as the top performer, delivering the highest accuracy when compared to the others. The system achieved an accuracy rate of 98% and demonstrated suitability for use across various lighting conditions.

The experiment aimed to evaluate the Bluetooth module's response time in locating the assistive cane under diverse conditions and distances from the object identifier and locator device. Table 3 illustrates the average response time of the Bluetooth module in various scenarios: an open environment (where the cane is placed without any obstruction), the cane covered with clothes (simulating a scenario where the cane might be concealed by fabric), and the cane placed inside a cupboard (representing an enclosed environment). The table represents the response time for distances of 3m, 5m, and 7m, tested across conditions of an open environment, covered with clothes, and in a cupboard. Each condition underwent 30 trials, and the average response time is presented with a margin of error of ±0.5m for distance measurement. The assessment of the Bluetooth module's response time encompassed diverse distances and conditions to locate the assistive cane. In an open environment, the module displayed swift responses: 2 seconds at 3m, 2 seconds at 5m, and 3 seconds at 7m.

Table 3: Results Of Bluetooth Module Average Response Time

Condition	Distance (in meters)	Average Response time (in seconds)	Number of trails
Open Environment	3 ± 0.5	2	30
Covered with clothes	3 ± 0.5	3	30
In cupboard	3 ± 0.5	4	30
Open Environment	5 ± 0.5	2	30
Covered with clothes	5 ± 0.5	4	30
In cupboard	5 ± 0.5	5	30
Open Environment	7 ± 0.5	3	30
Covered with clothes	7 ± 0.5	5	30
In cupboard	7 ± 0.5	6	30

When the cane was covered with clothes, response times slightly increased: 3 seconds at 3m, 4 seconds at 5m, and 5 seconds at 7m. Placing the cane inside a cupboard led to further delays: 4 seconds at 3m, 5 seconds at 5m, and 6 seconds at 7m. The module exhibited optimal efficiency in unobstructed environments but experienced notable delays in the presence of obstacles. These results establish a clear link between the Bluetooth module's response time, distance, and environmental conditions. Increasing the distance between the object locator and the assistive cane directly correlated with extended response times. Additionally, obstacles such as clothing or cupboard placement significantly interfered with signal reception, causing delays.

Promoting independence and autonomy for individuals with visual impairments is a key objective in assistive technology development. This study contributes to this goal by offering a detailed analysis in Figure 9 of how a blind individual utilized an object locator device over a week that is in a span of 7 days. The data precisely tracks the user's interaction frequency with this assistive technology for essential daily items, categorized by their placement at three distinct heights: floor, waist, and head levels. The items used for the observations are a charger, mobile, electric socket, cane, assistive cane (cane with Bluetooth).

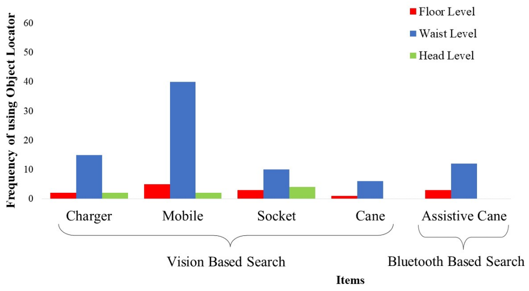

Figure 9 Usage Frequency Of Device Locator For Various Items At Different Levels Over 7 Days.

The findings present a clear picture of the user's consistent engagement with the object locator, with a particular focus on items positioned at waist level. The statistics reveals a marked preference for utilizing the device's Bluetooth functionality to locate the assistive cane, surpassing the frequency of traditional image processing methods. This comprehensive analysis transcends mere usage statistics, offering valuable insights into the user's behavioral patterns and preferences when accessing personal belongings. The emphasis on waist-level objects and the evident reliance on Bluetooth for the assistive cane illuminate the user's trust and dependence on this technology for navigating daily life.

While further investigation into long-term durability is essential, these results strongly advocate for the effectiveness and usability of this assistive technology, showcasing its potential to empower individuals with visual impairments to navigate their surroundings with confidence and independence.

5. CONCLUSION AND FUTURE WORK

This research has developed a Voice-Guided Indoor Assistance System tailored to enhance the independence and mobility of visually impaired individuals. This system embodies a multifaceted approach, incorporating innovative technologies to offer a comprehensive solution for indoor navigation and object retrieval. We meticulously curated a dataset, ensuring uniformity by resizing images and applying Prewitt edge detection. Feature extraction techniques, with ORB as the preferred choice, contributed significantly to the system's enhanced performance. The integration of Kmeans for clustering, along with classifiers like RF, DT, and K-NN, yielded 98% accuracy rate, highlighting the system's robustness across varying lighting conditions.

The hardware architecture seamlessly fused technology and user-centric design, introducing a novel two-way communication mechanism via Bluetooth between the user's primary device and assistive cane. This innovation not only aided in locating the cane or the developed device but also enhanced the user's autonomy and engagement, revolutionizing indoor navigation. The system's implementation, from user interaction through voice commands to object localization and precise navigation instructions, underscores its versatility. Its ability to handle multi-labelled images and prioritize object detection as per the user's request sets it apart as a valuable tool for the visually impaired. The inclusion of object localization ensures users can precisely locate and navigate essential objects.

In conclusion, this research represents a significant stride in assisting visually impaired individuals. Our model achieves remarkable accuracy and utility by thoughtfully integrating feature extraction, clustering, and classification techniques. Including object localization, voice commands, and Bluetooth-based interaction marks a novel contribution, enriching the lives of the visually impaired and exemplifying the potential of advanced technologies in addressing real-world challenges. As technology continues to evolve, our model serves as a testament to the transformative impact of thoughtful integration, enhancing the autonomy and engagement of visually impaired individuals in their surroundings. Future and enhancements include increasing the dataset size, incorporating natural language processing for more interactive feedback, utilizing more powerful hardware for complex scene descriptions, adding sensors for enhanced environment perception. For instance, further investigation into the long-term durability.

REFERENCES:

[1] Hussan, M. I. Thariq, D. Saidulu, P. T. Anitha, A. Manikandan, and P. Naresh, “Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People”, *IEEE Access*, Institute of Electrical and Electronics Engineers , USA, 2022, pp. 80-86.

[2] Benzarti, Faouzi, Hanen Jabnoun, and Amiri Hamid, "Object recognition for blind people based on features extraction", *International Conference on Intelligent Systems and Applications to Power Systems (IIPASC)*, Institute of Electrical and Electronics Engineers United States, 2014, pp. 1-6

[3] Mathew, Nikita Sara, Abraham Leo, Shebin Sam Sajan, Liza George, "VISION wearable speech-based feedback system for the visually impaired using computer vision", *4th International Conference on Trends in Electronics and Informatics* (ICOEI), Institute of Electrical and Electronics Engineers, USA, Vol. 48184, 2020, pp. 972-976.

[4] Islam, Raihan Bin, Samiha Akhter, Faria Iqbal, Md Saif Ur Rahman, and Riasat Khan, "Deep learning-based object detection and surrounding environment description for visually impaired people", *Heliyon* (Vol. 9, No. 6), Elsevier BV (Netherlands), June 2023.

[5] Badr, Georges, Kabalan Chaccour, "Computer vision guidance system for indoor navigation of visually impaired people", *2016 IEEE 8th International Conference on Intelligent Systems (IS)*, Institute of Electrical and Electronics Engineers , USA, 2016, pp. 449-454.

[6] Herperger, Miklós, Tibor Guzsvinecz, and Mostafa. "Indoor navigation for people with visual impairment using augmented reality markers", *2019 10th IEEE International Conference on Cognitive Informatics (CogInfoCom)*, Institute of Electrical and Electronics Engineers, USA, 2019, pp. 425-430.

[7] Xia, Yongquan, Yiqing Li, Jianhua Dong, and Shiyu Ma. "Research on Blind Obstacle Ranging based on Improved YOLOv5", *International Journal of Advanced Computer Science and Applications*, ISCA Technologies, Vol. 13, No. 10, India, 2022.

[8] Chavhan, Yashpalsing D., Nilakshi Mule, and Dipti D. Patil, "In-house object detection system for visually impaired", *International Journal of Future Generation Communication and Networking (IJFGCN)*, Vol. 13, No. 4, Foundation of Computer Science, USA, 2020, pp. 4919-4926.

[9] Al-Tourshi, Noura, Khuloud Al-Muqbali, and Al-Kiyumi. "Smart Technologies for Visually Impaired: Assisting and conquering infirmity of blind people using AI Technologies", *12th Annual Conference on Undergraduate Research on Applied Computing (URC 2020)*, Institute of Electrical and Electronics Engineers, UAE, 2020, pp. 1-4.

[10] Oleiwi, Bashra Kadhim, Farah F., and Muhsin Abdul M., "Real Time Blind People Assistive System Based on OpenCV", *Journal of University of Babylon for Engineering Science (JUBES)*, Vol. 28, University of Baghdad, Iraq, 2020, pp. 25-33.

[11] Devi, A., Therese M. Julie, and R. Sankar Ganesh, "Smart navigation guidance system for visually challenged people*", 2020 International Conference on Sustainable Engineering and Computing (ICSEC)*, Institute of Electrical and Electronics Engineers, USA, 2020, pp. 615-619.

[12] Simões, Walter C. S. S., Guido S. Machado, André M. A. Sales, Mateus M. de Lucena, Nasser Jazdi, and Vicente F. de Lucena, Jr., "A Review of Technologies and Techniques for Indoor Navigation Systems for the Visually Impaired", *Sensors*, Vol. 20, No. 14, MDPI, Switzerland, July 14, 2020.

[13] Rahmad, Cahya, Tanggon Kalbu Rawansyah, and Kohei Arai, "Object Detection System to Help Navigating Visual Impairments", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Science and Information (SAI) Organization, Vol. 10, no. 10, Germany, 2019.

[14] D. Plikynas et al., "Research advances of indoor navigation for blind people: A brief review of technological instrumentation", IEEE Instrumentation & Measurement Magazine, vol. 23, no. 4, 2020.

[15] Barontini, Federica, Manuel G. Catalano, Lucia Pallottino, Barbara Leporini, and Matteo Bianchi, "Integrating wearable haptics and obstacle avoidance for the visually impaired in indoor navigation: A user-centered approach" , IEEE Transactions on Haptics, vol. 14, no. 1, 2020.

[16] R. N. Kandalan and K. Namuduri, "Techniques for constructing indoor navigation systems for the visually impaired: A review", IEEE Transactions on Human-Machine Systems, Vol. 50, No. 6, 2020.

[17] Plikynas, Darius, Arūnas Žvironas, Andrius Budrionis, and Marius Gudauskis. "Indoor Navigation Systems for Visually Impaired Persons: Mapping the Features of Existing Technologies to User Needs", *Sensors 2020*, No. 3, , 2020, p.636.

[18] Nawaz, Waqas, Khalid Bashir, Kifayat Ullah Khan, Muhammad Anas, Hafiz Obaid, and Mughees Nadeem, "VIPEye: Architecture and

Prototype Implementation of Autonomous Mobility for Visually Impaired People", *International Journal of Advanced Computer Science and Applications,* Science and Information (SAI) Organization, Vol. 11, no. 5, 2020.

[19] Kassim, A. M., T. Yasuno, H. Suzuki, H. I. Sajini, S., & Pushpa, B., "Design Of Contemporary Multivariate Dataset to Assess The Quality of Object, Face and Proximity Detection in Assisting The Visually Impaired People", *Journal of Theoretical and Applied Information Technology(JATIT),* 101(23), 2023, pp. 7498-7510.

[20] Rusjdi, D., Heryadi, Y., Kusuma, G. P., & Abdurachman, E., "Estimating the Location of an Indoor Chair Object From a Single Image Using a Perspective Grid Approach" *Journal of Theoretical and Applied Information Technology(JATIT)*, 101(12), 2023, pp. 4884-4893.