# A Review on Voice Navigation System for the Visually Impaired

**BIMAL MOHAN[1], Ms. NAJLA NAZAR[2], EDWIN THOMAS[3], MUHAMMED ASHIK V A[4], VINAY M[5]**

*[1,3,4,5]Student, Department of computer science and engineering, Universal Engineering College, Kerala, India.*
*[2]Assistant Professor, Department of computer science and engineering, Universal Engineering College, Kerala, India.*

---***---

**Abstract:** A ground-breaking voice navigation system seems as a ray of light in a world where blind people must overcome enormous obstacles to navigate strange places. For those with vision impairments, safely and independently navigating in unfamiliar situations presents substantial obstacles. This system is a paradigm shift in providing safe and independent mobility assistance for the visually handicapped, as it makes use of cutting-edge technologies. The voice navigation system gives users the confidence to confidently navigate uncharted territory by seamlessly integrating computer vision and machine learning capabilities. With the use of real-time obstacle classification and recognition capabilities, the system can give users vital information about their surroundings, empowering them to make wise decisions while traveling. The system's smartphone-based platform, which provides an intuitive interface that can be accessed with voice commands, is its basis. By utilizing deep learning algorithms and pre-trained models, the system can precisely identify barriers and guide people across intricate environments. Precise positioning and route planning are made possible by the system's integration of GPS technology. Users receive up-to-date information on their location and the most efficient route to their destination via voice-activated instructions and advice. Individual user preferences are catered to by customizable options, which guarantee a tailored and user-friendly experience. A wider user base can utilize the system, democratizing access to necessary navigation tools, due to its affordability and interoperability with commonly available smartphone devices. The results of experimental tests highlight the effectiveness of the system, showing that it can reliably and accurately guide visually impaired people across a variety of surroundings. Through the promotion of increased autonomy and mobility, the voice navigation system enables users to move with unprecedented liberty and self-assurance, surmounting obstacles and opening up novel prospects.

***Key Words*: :** Voice navigation system, Machine learning, GPS technology, Deep learning algorithms, Navigation tools.

## 1.INTRODUCTION

Technologies designed to improve the lives of people with visual impairments have come a long way in the last few years. According to the World Health Organization (WHO), 36 million of the 1.3 billion people who live with vision impairments worldwide are considered to be blind. For those who are sight impaired, safely and independently navigating their surroundings is one of the biggest problems they confront. While they can provide tactile input, traditional aids like the white cane are not very good at giving specific information about barriers, like their kinds, distances, and possible dangers. This restriction highlights the pressing need for sophisticated navigation systems that provide all-encompassing environmental knowledge. Sophisticated technology like speech recognition, AI, and computer vision have been included into navigation systems designed for visually impaired users in answer to this demand. These systems enable users to move comfortably and independently by combining speech interfaces with object detection capabilities to deliver real-time advice and information. A key component of these navigation systems' object detection is computer vision technologies. Computer vision algorithms are able to recognize and categorize different types of barriers and environmental components by examining visual inputs from cameras or sensors. Convolutional neural networks (CNNs), one type of deep learning approach, have demonstrated impressive performance in object identification tasks, allowing navigation systems to identify objects, persons, and other pertinent aspects in the user's immediate proximity. Another essential element of contemporary navigation systems for the blind is voice awareness, which is made possible by speech recognition and natural language processing (NLP) technology. These systems give users real-time information about their surroundings, such as location-based notifications, obstacle alarms, and directional advice, through verbal feedback and audio signals. These navigation systems' precision and context awareness are further improved by the incorporation of GPS technology. Through the integration of GPS data with real-time object identification and voice interfaces, users can obtain precise navigational guidance that is customized to their specific location and surrounding circumstances. In order to create a comprehensive navigation system for people with visual impairments, this article investigates the combination of computer vision, artificial intelligence, speech recognition, and GPS technology. These systems seek to increase users' mobility, safety, and independence using a multidisciplinary approach and cutting-edge technical solutions, thereby improving their quality of life.

## 2. Object Recognition

Neural networks have transformed the area by enabling quick scanning of full images for multiple items, especially in object recognition techniques like YOLO (You Only Look Once). YOLO distinguishes itself as a more efficient and rapid substitute for previous detection techniques by segmenting images into a grid and forecasting bounding boxes with probability for every grid cell. It is perfect for real-time applications since it can evaluate the full image in one pass rather than continually scanning distinct image regions. YOLO improves accuracy even further by using anchor boxes to estimate bounding boxes[1]. Speech recognition, object detection, and image classification are just a few of the fields that have changed as a result of deep learning, and in particular, deep neural networks (DNNs). DNNs retrieve information directly from the data, as opposed to task-specific algorithms, creating more flexible and efficient systems[2]. A prominent deep learning technique called supervised learning improves system performance by using labeled data for learning. The system makes use of the MobileNet architecture in the field of item detection and recognition because of its lightweight design, high accuracy, and low computing overhead. The 21 layers that make up MobileNet's simplified architecture include both simple and deep convolutional layers. These layers use depth-wise separable convolutions to ensure processing efficiency[3]. This system's object recognition skills make a variety of activities easier, like helping those who are blind or visually challenged recognize familiar objects on the fly[4]. YOLOv8 is a better option than other methods for real-time multi-object recognition tasks because of its remarkable speed, accuracy, and efficiency. It is extremely useful in applications like augmented reality, surveillance systems, and autonomous cars because of its quick processing speed of many objects in real-time scenarios. All things considered, object identification systems have reached unprecedented heights thanks to the integration of neural networks, especially YOLO and MobileNet, which enable precise and efficient object detection in a variety of settings and circumstances. These technologies show promise for future breakthroughs across a range of industries and the creation of creative answers to pressing problems in the real world as they grow further.

## 3. Voice Navigation

Within the field of accessibility technology, voice navigation systems are innovative leaders that provide the visually impaired people with a means of empowerment and freedom. For people who are visually impaired, these solutions bridge the gap between the digital and physical worlds by utilizing the power of modern voice technology. The complex interactions between text-to-speech (TTS) and speech-to-text (STT) technologies provide the foundation of a voice navigation system. The process starts with textual instructions being converted into clear voice using powerful TTS engines such as Espeak[5]. This open-source miracle is a multipurpose tool for converting textual material into immersive audio experiences, in addition to unlocking a multitude of languages. With the help of Espeak, users may navigate through intricate environments with a variety of landmarks, like bus stations, colleges, hospitals, and more, all while receiving clear voice instructions. Concurrently, the system embraces the dynamic nature of voice recognition, as exemplified by tools like the Google Voice Recognition API and Pocket Sphinx. By taking spoken inquiries from users and turning them into text that can be used, these engines serve as gateways. With its trainable capabilities and configurable dictionary, Pocket Sphinx assumes the role of user input deciphering, but with some potential for accuracy improvement. But the arrival of the Google Voice Recognition API signals the beginning of a new age in accuracy, providing unmatched precision and adaptability in comprehending user commands. Users' vocal instructions are effortlessly understood and returned in a symphony of interaction that results from the synergy between TTS and STT. This harmony goes beyond simple direction finding; it includes a comprehensive network of support, from locating particular data, such addresses, to enabling natural interactions with the digital world. The importance of voice navigation systems extends beyond their technological features to include their societal impact. With an estimated 253 million people worldwide suffering from vision impairments, these technologies become important tools for accessibility and inclusion[6]. They break down barriers, enabling people to confidently move across real spaces and enjoy the digital world with a fresh sense of independence. In this age of swift technology advancement, voice navigation systems serve as iconic examples of the revolutionary potential of invention. They represent the fusion of human-centered design and state-of-the-art technology, providing the visually impaired people with a route to inclusivity and empowerment in addition to navigation.

## 4. Distance Computation

An essential component of computer vision is the distance computation technique shown here, especially when it comes to comprehending the spatial relationship between a camera and objects in its field of view. The per-frame triangulation geometric methodology, which estimates distances based on known parameters and image attributes, is leveraged by this method. Fundamentally, the technique depends on the camera lens's focal length, the object's actual height, the camera frame height, the picture height, and the sensor height. In order to convert image data into real-world spatial measurements, these characteristics are essential to the calculating procedure. Let's start by discussing the idea of per-frame triangulation. Triangulation is a geometric technique that measures the angles between two known locations to find a point's location in space. Per-frame triangulation in computer vision refers to the process of determining an object's distance from a camera by utilizing data from a single image frame. Per-frame triangulation treats each frame separately, in contrast to techniques that depend on temporal information or consecutive frames. This makes it appropriate for real-time applications where the processing of incoming data must occur continuously. The camera image system's critical components are the important parameters that are used in the distance calculating procedure. One essential component that establishes the camera's field of vision and influences the perspective of the picture it captures is the focal length of the lens, represented by the symbol F. It directly affects the scale of things in the image and is commonly measured in millimeters. While a lower focal length offers a larger field of view with less magnification, a longer focal length narrows the field of view and amplifies distant objects. The object's actual height, represented by Rh, is a measure of the object's true size in the real world. This parameter, which is commonly expressed in millimeters, is used to determine the scale of the object in the taken image. Since the object's height directly affects the triangulation process, accurate measurement of the object's height is necessary for accurate distance calculation. The height of the camera frame, defined as $Fh$, the term "Fh" describes the height in pixels of the camera frame. This parameter, which determines the vertical resolution of the acquired image, is a feature of the image sensor. A higher frame height makes the acquired image more precise and detailed, which improves distance estimate. The physical dimensions of the image sensor, which are commonly expressed in millimeters, are not the same as the camera frame height. Similarly, the vertical resolution of the acquired image is represented in pixels by the image height, which is represented by the symbol Ih. It is an essential triangulation process parameter that is used to scale objects inside the image. The scale of the scene that is captured is determined by the relationship between the image height and the camera frame height, which also has a direct impact on the distance estimation accuracy. The sensor height, represented in millimeters by the symbol Is, is the camera sensor's actual size. It immediately affects how the collected image is projected onto the sensor plane and is an inherent feature of the camera system. The focal length of the lens and the sensor height together define the physical characteristics of the imaging system and have an impact on perspective distortion. The relationship between these crucial factors is summarized by the formula D=F×Rf×Fh/Ih×Sh, which gives an estimate of the distance from the camera to the identified item[7]. The dimensions of the object that the camera captured are represented by the numerator of the formula, while the dimensions of the image and the camera sensor are represented by the denominator. A ratio is produced by dividing the numerator by the denominator, which connects the object's actual dimensions to how it is shown in the picture. With the help of this ratio, which acts as a scaling factor to translate pixel measurements into actual distances, precise distance estimation based on the size of the item in the image and the camera's settings is possible. The distance computation approach utilizes the geometric attributes of the camera system and per-frame triangulation to accurately localize objects within the camera's field of vision. To sum up, the distance computation approach shown here is a core component of computer vision and allows one to estimate the distances between objects in a camera's field of view. Through the use of per-frame triangulation and critical variables including the camera lens's focal length, the object's actual height, and the image and sensor's dimensions, the technique offers a precise and dependable way to localize objects spatially in real-world settings. Approaches like this will become more and more crucial in a variety of applications, from augmented reality to autonomous navigation, as computer vision technology develops.

## 5. CONCLUSIONS

Voice navigation systems are unique in that they are powerful instruments that make use of cutting-edge speech-to-text and text-to-speech technologies to improve user engagement with both the digital and physical worlds and to deliver clear directions. These systems greatly advance accessibility and societal inclusion in addition to encouraging independence. For the purpose of precisely measuring the distances between objects in a camera's field of view, the distance computing technique through per-frame triangulation is essential in the field of computer vision. Precise distance estimate is made possible by the interaction of characteristics like focal length, object height, camera frame height, image height, and sensor height, which convert picture data into real-world spatial measurements. Furthermore, the capacity to identify and recognize items accurately and effectively has been revolutionized by neural network-powered object identification algorithms like YOLO and MobileNet. These technologies have several uses, ranging from improving real-time tasks like surveillance and autonomous navigation to helping visually impaired people recognize objects. In summary, the convergence of cutting-edge technologies such as neural network-based object recognition, computer vision techniques for distance computation, and voice navigation systems has greatly improved accessibility and efficiency across a range of domains, indicating potential for future innovation and impact.

## ACKNOWLEDGEMENT

The authors extend sincere gratitude to Mrs. Najla Nazar for her Guidance and project coordinators, Mrs. Najla Nazar and Mrs. Gishma K M for mentorship. Their expertise played a vital role in shaping the direction and focus of this research.

## REFERENCES

[1] Abdul Muhsin M, Farah F. Alkhalid, Bashra Kadhim Oleiwi, ”Online Blind Assistive System using Object Recognition”, 2019.

[2] Fazal Wahab1 Inam Ullah, Anwar Shah, Rehan Ali Khan, Ahyoung Choi, Muhammad Shahid Anwar, ”Design and implementation of real-time object detection system based on single-shoot detector and OpenCV ”, 2022.

[3] Fahad Ashiq, Muhammed Asif, Maaz Bin Ahmad, Sadia Zafar, Khalid Ma sood, Toqeer Mahmood, Muhammad Tariq Mahmood, Ik Hyun Lee, ”CNN Based Object Recognition and Tracking System to Assist Visually Impaired People”, 2022.

[4] Jamil Abedalrahim Jamil Alsyayadeh Hasvinii Baskaran, Irianto, Maslan Zainon,”Intelligent Interfaces for Assisting Blind People using Object Recognition Methods”, 2022.

[5] Anushree Harsur and Chitra.M , Asst.prof. Bangalore, India, ”Voice Based Navigation System for Blind People Using Ultrasonic Sensor”. IJRITCC, June 2015.

[6] Harsha Gawari and Prof. Meeta Bakuli, ” Voice and GPS Based Navigation System For Visually Impaired”. Harsha Gawari et al Int. Journal of Engineering Research and Applications, April 2014.

[7] Saleh Shadi, Saleh Hadi, Mohammad Amin Nazari and Wolfram Hardt, “Outdoor Navigation for Visually Impaired based on Deep Learning”, Nov 2019.