A Mobile Augmented Reality Assistant via Deep Learning and Lidar for the Visually Impaired and Blind

by

Tianshi Xie

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 06, 2023

Keywords: Augmented Reality (AR), Visually Impaired, Deep Learning, Mobile Computing, Computational Speed, Intelligent System

Approved by

Cheryl D. Sealsl, Chair, Charles W. Barkley Professor of Computer science and Software Engineering
Tao Shu, Associate Professor of Computer science and Software Engineering
Wei-Shinn (Jeff) Ku, Professor of Computer science and Software Engineering
Dallin Bailey, Associate Professor of Speech Language and Hearing Sciences

Abstract

Visually impaired and blind (VIB) people often face additional difficulties in their daily lives due to their lack of access to visual information, which reduces their quality of life. Due to their vision problems, it is hard to avoid obstacles and find target objects in different spaces. Assistive devices have been developed to help blind people avoid obstacles and navigate. However, many of these technologies require users to purchase additional devices, and they need more flexibility, thus inconveniencing VIB users. To address the above issues, we proposed a new approach, implementing a design, programming, and interface to create a Navigation Assistance through AR technology and Deep learning (NAAD) system. The NAAD system is based on Mobile-Net Single-shot Detection (MobileNet-SSD), Augmented Reality (AR), and LiDAR for implementing obstacle detection, target object detection, distance calculation, and navigation of user-specified lost objects. It includes a mobile application that uses simple voice and gesture controls to aid navigation. This system aims to (i) help visually impaired and blind (VIB) people avoid obstacles in daily life, (ii) Use computer vision to find user-specified objects quickly, and (iii) integrate LiDAR for AR/VR experiences, reducing additional equipment and improving the distance accuracy between the obstacle and the user. In my research based on the NAAD system, we designed the safe mode and query mode. Subsequently, We upgraded the system and added the navigation function in the query mode.

In the first study in this dissertation, we discussed object detection integration and introduced the query mode of the NAAD system. The object detection function of the NAAD system is implemented by machine learning. This virtual assistant provides different functions such as obstacle detection, distance estimation, navigation system, and real-time environment analysis to help users detect and find their object items. Subsequently, the object detection feature includes a unique interactable feature that enables the user to interact with the device to find indoor objects by providing voice and vibration feedback and further provides voice navigation to the user. Experimental results show that the system has a good performance of the response

time of 19 ms outpacing [44], and the FPS is over 30 frames per second and outperforms similar systems [34] and [23].

In the second study in this dissertation, we discussed obstacle detection integration and introduced the safe mode of the NAAD system. In safe mode, this virtual assistant provides different functions such as obstacle detection, distance estimation, and an alarm system that analyzes the environment in real time and alerts the user to avoid obstacles. To improve the distance detection accuracy between obstacles and users, we introduced a LiDAR sensor instead of a depth camera to actively detect the environment without being affected by ambient light. The experimental results show that the distance detection accuracy between the obstacle and the user is 96% within the five-meter range, outpacing other research and surpassing similar projects.

In the third study in this dissertation, we optimize and upgrade the NAAD system. We discuss object-based navigation and introduce an object navigator in the query mode of the NAAD system. We designed the object navigation module by adding a memory storage unit that records the target object and the user's location in real-time. It can solve the detection and navigation of the target object when the user is far away from it or not in the same space. Finally, we introduce the frame and interface design of the upgraded final NAAD (Navigation Assistance through AR technology and Deep learning) system. Experimental results show that the NAAD system can successfully locate and navigate to lost item in different spaces.

Acknowledgments

Throughout my entire Ph.D. path, I have received enormous advice and support from many people in various ways. It is my great pleasure to express my infinite gratitude and sincerest gratitude to everyone who has helped and encouraged me over the past four years.

First and foremost, I would like to give my sincere gratitude to my advisor, Dr. Cheryl Sears, for being an incredible doctoral supervisor and a fantastic teacher for providing me with a warm and outstanding intellectual environment at Auburn University. She gave me the freedom to explore independently and offered me helpful guidance and constant encouragement when I stumbled. I appreciate all her time, ideas, and funding that made my Ph.D. thesis possible. Also, I would like to thank all my committee members: Dr. Cheryl Seals, Dr. Tao Shu, Dr. Ku, and Dr. Bailey, for their valuable feedback, support, and guidance.

Many thanks go to the members of the AR/VR research group: Sathish Akula, Alexicia Richardson, Majdi lusta, Sean Kim, and Nikolay Sargsyan for many exciting and inspiring discussions.

In addition, I would like to acknowledge my internship manager, Harpreet Singh of Amazon, for his trust in my programming abilities. This awesome internship taught me how to apply what I learned in school to a real-world business environment, significantly improved my software programming skills, and expanded my mastery of research techniques.

I am also grateful to my parents and my family for all their love, dedication, and encouragement. Finally, a special thanks to my wife, Ting, for being supportive with unconditional love and embracing me.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

According to the World Health Organization (WHO), there are more than 285 million visually impaired and blind (VIB) people worldwide at present. The growth rate of the number of VIB people is very high. This number is expected to triple in the next 30 years. In the United States, approximately 12 million people aged 40 years and older have vision impairment, of which 1 million are blind, 3 million have vision impairment after correction, and 8 million have vision impairment due to uncorrected refractive errors. Currently, there are 4.2 million Americans age 40 and older with uncorrectable visual impairments, 1.02 million of whom are blind. This number is expected to increase by 2050 due to the growing prevalence of diabetes and other chronic diseases, and the rapidly aging U.S. population more than doubled to 8.96 million. The proportion of children under the age of 18 years with eye and vision problems is approximately 6.8% in the United States. Nearly 3% of children under 18 years old are blind or have impaired vision, even with glasses or contact lenses. According to the National Institute for Occupational Safety and Health (NIOSH), approximately 2,000 U.S. workers each day require treatment for work-related eye injuries. In the National Eye Health Education Program (JHEP) 2005 Public Knowledge, Attitudes, more than 70% of respondents consider that their vision loss would significantly impact their daily lives.

Visually Impaired and Blind (VIB) because they lack the ability to perceive visual information and cannot accurately judge changes in the surrounding environment, which limits their ability to process their surroundings and interact with society, hinder their daily activities, and increase indoor and outdoor occurrences risk of an accident. They often face many challenges in their daily lives, and finding lost items indoors is one of the biggest challenges faced by

the visually impaired. Due to their vision problems, it can be hard to find the target object, especially when they are in a different space from the target item. They usually spend more time and energy than healthy people. These can cause them to have high levels of anxiety and depression. Therefore, VIB people with reduced mobility are most in need of assistance. Using assistive technology to make their lives better and more accessible is significant. Traditional assistive devices such as guide dogs, walking canes, and glasses have been used by the visually impaired and blind (VIB) to perform basic daily tasks. However, while guide dogs can detect obstacles, their communication with human family members is often unclear, and they cannot help the VIB people to find the target object. Walking canes are very effective in detecting ground anomalies, but the use of walking canes requires constant detection and is not effective in detecting objects above the waist. Glasses can only help people with mild visual impairment. They cannot do anything for severely visually impaired or blind (VIB) people. As technology has evolved in recent years, the development of assistive technologies for VIB people has gradually increased. Many assistive navigation devices have been designed to assist VIB people in real life, such as Depth Cameras, Radio Frequency (RF), Ultrasonic Sensors, and Infrared Sensors. But these devices need to be purchased additionally and lack flexibility, which is inconvenient for VIB users. To address the above issues, we designed an application named NAAD for LiDAR-enabled mobile devices. It can improve the quality of life for the VIB people by enabling obstacle avoidance, object detection, and navigation of user-specified lost items through machine learning and AR technology.

In this study, we will focus on the research, design, and development of three functions of the NAAD system: (1) object detection, (2) obstacle detection and avoidance, and (3) target object navigation. This dissertation consists of four parts. Specifically, in the first part, we discuss object detection techniques. According to this research result, the Query Mode of the NAAD system was designed. In the second part, our work focuses on obstacle detection techniques for obstacle avoidance. On this basis, we designed the Safe Mode of the NAAD system. In the third part, we discuss indoor navigation techniques for target objects. Then we developed a navigation module with a new approach to help VIB people navigate the missing

target objects in different rooms. The fourth part introduces the overall design and experiments of the NAAD system.

Chapter 2

Literature Review

2.1 Augmented Reality technology to assist the VIB people

Visually impaired and blind (VIB) people can not read the door number when they reach their destination. They do not know which door they should enter. Therefore, Jonathan Huang [22] utilizes AR technology to enhance the user's visual information and help them quickly obtain information. The project they proposed is based on Microsoft HoloLens. This device integrates augmented reality, optical character recognition, OCR control, Voice Control, Wi-fi, and Bluetooth Technologies. The software development mainly uses Unity Technologies and Microsoft's HoloLens toolkit. First, the user uses the virtual cursor provided by the device to identify the target in the virtual area in front of the user. The identified target will be overlaid by a 3D model called the AR sign. The user can move the cursor to select or cancel multiple AR signs in front of them. When pressing the AR SIGN, the user can read the information through a Manual touch button or use a voice command called "show me. ". Then, this information will be magnified in front of the user and read aloud by the system. In the experiment, they divided participants with the same visual impairment into two groups, the AR group, and the control group. The results showed that the AR Group scored significantly higher than the control group. In addition, they also analyzed the trajectories of the two groups of participants in the experiment. The AR group takes fewer paths than the control group because the AR group can use the AR device to remotely scan the target's room number, while the control group has to walk into the target to read the data. However, the AR Group takes longer to complete the task than the control group because it took a certain amount of time from the device scan information

to the resulting feedback. The system can also help people with visual impairments to read information about indoor navigation. It has got the user's approval. Through experiments, we found that AR technology has great potential in enhancing visual information. This technology can improve people's lives and can also be embedded in other mobile devices.

HyeongYeop Kang [28] uses AR technology to analyze real-time user environments and remind users to avoid obstacles. The system has an obstacle detection module and a direction guidance module. The obstacle detection module uses Google AR-Core to scan the surrounding environment and calculate the vertical height difference of all feature points. If the height difference exceeds the threshold A = 10 cm set by the system, the characteristic point is the point on the obstacle. And if the number of characteristic points of an obstacle exceeds another threshold B = 3. The system will issue a warning. For the direction guidance model, the system first defines the Roi of the region of interest. The system only scans the feature points in the region of interest, and the number of feature points is at least 20. Second, the user must hold the phone at an angle between 30-60 degrees, or the system will give a hint. The distance between the last frame of the video captured by the camera and the current frame is at least 0.1 s. In the experiment, the authors developed a simple AR game while turning on the obstacle monitoring system and letting the participants play for 10 minutes. Obstacles detected fall into four categories: obstacles, walls, staircases, walls, and curbs. Finally, six indexes of the system are evaluated by Friedman Test: Visual - Audio, Visual - Tactile, Visual -No, Audio-Tactile, Audio-No, and Tactile-No. It turns out that auditory and tactile interfaces are two of the user's favorite reminders. The system proposed in this paper can effectively help AR users avoid obstacles in front. But the system still has some drawbacks. For example, the performance of the system will be affected by light. Especially at night or in dark places, it is difficult for the system to identify obstacles at characteristic points. In addition, it is not easy to use the device to a certain extent.

2.2 Deep learning technology to assist the VIB people

Traditional methods such as walking canes and guide dogs cannot identify the type of obstacle and may lead to unnecessary accidents. Researchers take advantage of the depth camera,

radiofrequency, ultrasonic sensor, or infrared sensor to develop many navigation tools to help the VIB peoplepeople with these two problems. Such as Etas, EOAs, and PLDs. However, one problem these traditional tools have in common is that they can only identify very few types of obstacles. Bor-Shing [24] Lin proposes a system based on deep learning to solve this problem. The system only requires the user to have a mobile phone without taking any other detection equipment, and the system can work indoors and outdoors. This system has two modes: Online Mode and Offline Mode. It can automatically switch modes according to the WI-FI situation. In Online Mode, the system provides a stable mode and a fast mode according to the user's settings. In Stable Mode, TCP is the network connection protocol between the client and the server, and Faster-R-CNN is used in the image recognition module of the server. In Fast Mode, the network protocol is RTP, and Yolo is used as the image recognition module on the server. In the direction and distance module, the author uses the Mono-SLAM formula to calculate the distance between the camera and the target. The distance can be divided into three grades: close, medium, and far. Then split the image into three horizontal sections marked left, center, and right. Each segment represents the direction of the object. When the server completes the image processing, the system will output the category of the target object, the distance between the target object and the user, and the direction of the target, and send the final information to the user. The experimental results verify that mixing the public image data set Pascal with the images taken by the author can improve the recognition rate of the model. Secondly, if the detected obstacle is more than 10 meters away from the user, the error will increase. Third, User satisfaction with the system reaches more than 60% including overall impression, user interface and user experience, and alert frequency. This paper mainly develops mobile navigation software for people with visual impairment. The obstacle recognition degree, computation performance, and portability of the system equipment of this system have Greatly enhanced over traditional tools. In addition, the performance of the system will be further improved through the hardware upgrade of the server and the improvement of the data set quality.

Soobin Ou [32] proposed a system, which uses the image recognition function of artificial intelligence to analyze the environment around the user and then guides the user to avoid obstacles by voice guidance. The system architecture is divided into two parts: Client and

server. In the client, the system gets the video of the user's environment in real time through Raspberry Pi and sends it to the server for data analysis. The client will broadcast the results returned by the server in the corresponding voice. In the server, the system first preprocesses the video stream from the client to a sequence of 30 frames per second, then passes each image in turn to the obstacle recognition network. If the network output shows Bollard for the obstacle in this picture, the result is given to the Location Calculation Module. This module will send the location of this obstacle to the client if the obstacles in the picture are identified as a crosswalk light walk. The system will pass this information to the Crosswalk Discrimination module. This module outputs the status of the Crosswalk Light: None, red, flashing, or green. And send it to the client. The client will help the user through the intersection based on the status of the crosswalk light. This system calls two kinds of neural networks separately and carries on the performance comparison. The results show that Faster R-CNN is better than SSD Mobile-Net in speed and accuracy. As artificial intelligence continues to evolve in computer vision, researchers spend a lot of time on more popular missions, such as object classification, detection, and segmentation. But there is little research on distance estimation, which is the estimation of the distance between the camera and the target object. In autonomous driving, the vehicle can adjust its speed and direction to avoid collision with other vehicles according to the estimated distance. The traditional methods include inverse perspective mapping (IPM) and support vector regression (SVR). The performance of these algorithms is not good enough, especially when estimating the distance of objects over 40 meters or on either side of the camera. Then the author proposed a basic model and an enhanced model to solve this problem. Both models use deep learning techniques. The experimental results show that they are better than the traditional algorithms, and the enhanced model is the best.

Zhu Jing [25] first proposed a basic model consisting of three modules: feature extractor, distance regression composer, and multi-class classifier. For the feature extractor, the author uses vgg16 and Res50 as the feature extractor. The extracted feature vector is then passed into ROI along with the object's bounding box. ROI is a pooling layer. ROI produces a fixed-size feature vector Fi to represent each object in the image. Use the results generated from Roi as input data for Distance regression and Classifier modules. The Distance and Classifier modules

will output the distance from the camera to the target object and the category of the target object. The final loss of the base model will be calculated by the loss of the distance module and the classifier module. For the enhanced model, the only difference from the base model is a key point regressor module. This module can improve the accuracy of estimating distance. The experimental results show that the proposed model is more accurate than SVR and IPM in distance estimation and target object classification. It also successfully predicted the distance of objects over 40 meters or on either side of the camera. The proposed model solves these two challenging problems successfully. Moreover, Kitti and Nuscenes' experimental data were reconstructed with the LiDAR point cloud, which solved the missing data problem. This neural network model can use in robotic systems and autonomous driving systems. It can also use in other fields, such as navigation systems for the disabled.

Obstacle avoidance has always been a hot topic in robotics, but it also exists in autonomous driving and blind guidance. Especially identifying small obstacles in these areas is a difficult point. There should be a few solutions to this difficulty at present. It includes the Range-based method, appearance-based approach, and semantic segmentation only with RGB information. None of them can detect small obstacles. To solve this problem, Minjie Hua [26] proposes a method based on RGB-D semantic segmentation, which can effectively detect small obstacles indoors or outdoors. This system consists of three modules: RGB-D based on two-stage Semantic Segmentation, morphological processing, located destination setting, and path planning. In the RGB-D module, the author uses ResNet-50 to extract the feature map and then extracts the ROI-processed contour map by Red-net, then passes the contour map To ResNet-34, and finally outputs the classification result: Road, obstacle, and others. In the Morphological Processing module, the system performs an erosion operation and then a dilation operation, which effectively removes noise from the image and improves data quality. For the local Destination Setting, the author uses a progressive scanning method to scan the image line by line, Pixel by Pixel for each line, and ultimately determine the point of the destination. For Path Planning, the system generates a path based on the target point obtained in the previous step and uses the Artificial Potential Field (APF) method. The robot adjusts the direction and moves according to each point in the path until reaching the target point. The experimental results

show that the performance of the proposed semantic analysis network model is better than that of other network models including Seg-Net, Fuse-Net, and Red-Net. Besides, the accuracy of the system-generated path is much higher than that of the Stereo method. The system proposed in this paper can effectively help the robot to avoid small obstacles. To improve the accuracy of this system, the author adopted a series of methods, including using Flow-Net to adjust the resolution of the feature map, denoising the generated semantic map, and preprocessing the blurred image caused by insufficient light.

Jason, Jiang [27] presents a method based on the ALEXNET neural network to help robots avoid obstacles effectively. To solve the influence of illumination on image recognition, the author took a depth map as training data and compared it with an RGB image only. The results show that the depth map can effectively overcome the impact of illumination on image recognition. And in the case of insufficient night light, the accuracy of the depth map is better than using only RGB images. The system uses Intel's real-sensing depth camera D435. The D435 integrated Stereo vision and structured light technologies. It can record video at 6kph and generate depth images and RGB images. The two images are input into the ALEXNET model as training data, respectively. Because the training time of the traditional ALEXNET model is too long, the author adopted the Transfer learning technique to replace the last layer of Alex-Net into three fully connected layers, and only train the last three layers. This operation can save lots of training time and data sampling time. The experimental results show that the accuracy rate of the model based on the depth image during the day is 80.51%, while the accuracy rate of the model based on the RGB image is 80.53%. There is not much difference between them. At night, the accuracy rate of the model based on the depth image is 85.42%, while the accuracy rate of the model based on the RGB image is 71.07%. The authors found that the performance of the depth-map-based model did not perform as well during the day as it did at night. The main reason is that this depth camera integrates an infrared function in the case of sufficient sunlight, the generated depth map will produce noise, which will affect the accuracy of image recognition. So, the authors propose to make up for this deficiency by using the two models alternately during the day and at night. With the help of a depth map, this system solved the influence of illumination conditions on image recognition and achieved

good results. But this method is only suitable for robots of fixed height. That is, when the height of the robot changes, the perspective of the image changes. It will cause the training data to have to be re-collected and the model to be re-trained. Biological binocular vision can provide depth information of the image. The paper used the disparity map to restore the depth information of the images according to the difference between the two cameras. According to the Computer Vision Triangle Algorithm, the depth is converted into the actual distance between the target and the camera. It can effectively make the robot avoid obstacles on the road. In addition, semantic analysis can be used to identify roads and obstacles in the scene. Finally, the Hough algorithm is used to make the robot walk on the right side of the road which can prevent the robot from entering an unknown area. The experiments show that the robot can reach its destination safely through the distance estimate module, the semantic analysis module, and the Hough module.

H. Chen [23] presents a system that is mainly composed of two neural networks. They are the Vision Disparity Network and the Semantic Segmentation Network. To the vision Disparity network, first, predict the Disparity maps using the image on the left. Then generate the target image left and right according to the Disparity maps and original images. Finally, reducing the difference between the generated image and the original image can improve the accuracy of the dissimilarity Map. The estimated dissimilarity map is then passed into a Back Propagation Neural Network BPNN and trained according to the transformation formula from vision disparity to depth. At the same time, the author uses ENET as a Semantic Segmentation network to analyze the whole scene. Finally, more than 500 pixels of obstacles in the image are filtered out. According to the distribution of the obstacle depth value in the histogram, consider the obstacle depth value closest to the camera as the effective depth distance. The system proposed in this paper successfully helps the robot recognize short-distance obstacles and reach the destination during navigation. In addition, the system is very scalable. The obstacle depth screening algorithm can modify to classify multiple obstacles into three different levels: high, medium, and low according to distance. The robot can formulate different direction strategies according to different levels of obstacle information to achieve the effect of dynamic optimization of the path.

A. K. Srinivasan [34] aims to help visually impaired people avoid obstacles indoors. The system mainly uses stereo vision and neural networks to help visually impaired people identify surrounding objects and obstacles. And according to the distance detection system, to effectively avoid obstacles and reach the target object. The experiment shows that this method can effectively help the visually impaired locate the target object and provide sound feedback to the target object and the obstacle. The system consists of three modules: Object detection, depth calculation, and text-to-speech. For the Object detection module, the authors used the SSDLIFE MobilenetV2 model because it was Faster and more accurate than popular algorithms such as Fast R-CNN, Faster R-CNN, Yolo, and Fast Yolo. The input data for this model is the video frame, and the output data is the item category and the bounding box of the item. The training data used in the MS COCO dataset contains 90 different categories of items. For the depth calculation, the author uses Stereo Vision To calculate the distance based on camera parameters and disparity. The disparity is obtained from the center point of the object bounding box in the left and right images. Finally, the author classifies objects and obstacles based on the fuzzy algorithm and sends the object and obstacle information to the user through the Text to Speech engine. Experiments have shown that too high a frame rate will increase the amount of calculation, so they adjusted the frame rate to 15 FPS. In addition, when the two cameras are 4 inches apart and the object depth is 10 to 50 inches, the system can achieve good results. Proposed in this article can effectively help the visually impaired recognize surrounding objects or obstacles. But the system still has some limitations. First of all, the author did not provide a solution to the light effect. Second, changes in the height of the equipment will have an impact on the results. Finally, the device is not convenient to carry.

Yimin Lin [33] uses neural networks (CNN) to identify obstacles in the environment of visually impaired people and conduct safe navigation. Besides, a semantic analysis neural network (Fuse-Net) is used to generate semantic maps of the environment of the visually impaired. According to the generated semantic map, visually impaired people can effectively extract the information they are interested in. Experimental results show that the system can effectively help the visually impaired avoid obstacles and reach their destination quickly. The system hardware includes an RGBD camera, headset, processor, and touch interface. The software part of

the system has two modules: the navigation prediction module and the Semantic analysis module. For the navigation prediction module, the author uses the Google-Net neural network to predict the navigation guidance. The RGB, DEPTH, and semantic graphs are used as input data. The output includes a left turn, a right turn, and straight-line feedback to the user in the form of sound. For the semantic analysis module, users can click on the generated semantic map at any location to obtain the corresponding recognition information and feedback to users in the form of voice. The author uses an RGBD camera to collect training data, which contains 15,000 and 6,000 indoor and outdoor images. The experimental results show that the precision of the navigation prediction module can reach 99.6% and 98.2% day and night. Users rated the system's four metrics: Navigation, Traffic rounding, Object searching, and Layout parsing. Compared with traditional methods, user satisfaction has increased by more than 60%. The system can help the visually impaired identify the surrounding environment and guide them to reach their destination safely and quickly. The performance of the system and the user experience have achieved good results. Besides, the system still has a lot of room for improvement. For example, add other obstacle detectors, optimize obstacle recognition neural networks, and improve GPU and other hardware devices.

2.3 Indoor Navigation technology to assist the VIB people

Visually impaired people face the challenge of orientation and mobility in outdoor and indoor scenarios. Specifically, the issue of indoor navigation is a difficult one as satellite signals from the Global Positioning System (GPS) do not easily penetrate inside buildings. Current indoor navigation systems tend to have drawbacks, such as requiring complex installations or specific conditions that are hard to encounter in real life or relying on positioning technologies that have limitations in accuracy.

Martinez-Sala [1] presents SUGAR. This indoor navigation system has been designed to help visually impaired people navigate through buildings with high accuracy, robustness, and ease of use. It can provide a long-range, high-precision positioning technology that can operate in the presence of obstacles, and multipath signal reflections are highlighted as a solution to improve the performance of indoor navigation systems for visually impaired people. The

system uses Ultra-Wideband (UWB) positioning techniques in a two-dimensional space to locate the user and a spatial database created from a digitalization of the floor plan of a building to guide the user to their destination or through a route with multiple points of interest. The system applies the A* algorithm to the information in the database to find the appropriate path and guides the user through voice commands and acoustic signals. The user interacts with the system through acoustic signals and voice commands played through headphones. The system is simple to install and is suitable for large public buildings, but it can also be deployed in other scenarios.

Madoka-Nakajima [2] proposed a new indoor navigation system that utilizes visible light communication technology and a geomagnetic correction method to assist visually impaired individuals in navigating indoors. This system aims to address the challenges faced by visually impaired people in navigating indoors, such as the difficulties in acquiring accurate positional information and detecting directions. An experiment was conducted with visually impaired participants to evaluate the effectiveness of the system, and it was found that the system was able to provide accurate positional information and travel direction using LED lights and a smartphone's geomagnetic sensor. The system utilizes LED lights to transmit visible light IDs, which allows for accurate indoor positioning by measuring the azimuth based on the true north and identifying the geomagnetic distortion. This technology is less affected by environmental factors such as obstacles, which makes it more reliable than other indoor positioning methods. Additionally, it uses a correction algorithm with the geomagnetism beneath the LED lights, which allows for more accurate travel direction guidance. The system was tested with visually impaired people, and the results were positive, where the subjects were able to reach their destination with the help of the system. However, there are still some deficiencies in the design of the system. First, Since it is difficult to hold a smartphone still while holding a cane, the authors propose that this problem could be solved by attaching an optical sensor to the belt, but this would add additional equipment and reduce the ease of use of the system. Second, the Timing of spoken navigation does not match the traveling speed of visually impaired people who walk fast. Third, blind people's perception of whether they are on the right track can be resolved by providing continuous feedback sound during travel. However, such continuous

sound feedback may also be regarded as useless information by the user and reduce the user experience of the user.

Serdar Bilgi [3] proposed an indoor navigation system for hearing and visually impaired people called Loud Steps at Istanbul Technical University (ITU) in Turkey. The system uses voice features to guide users to various destinations, such as offices, homes, schools, hospitals, and public institutions. It has been downloaded and used by 130,000 users in 11 countries, including over 20,000 visually impaired people. The system was first implemented in the Faculty of Civil Engineering at ITU's Ayazaga Campus and has since been established in four additional buildings on campus. The project plans to establish the system in the remaining 12 faculties and six institutes on campus and to integrate an outdoor navigation system for the entire 2.5 million square meter campus area. The system provides both voice and visual guidance to users. The Loud Steps project is a self-help application designed to aid visually and hearing-impaired people in navigating closed areas without assistance. The app is available on iOS and Android operating systems and is used by more than 130,000 people in 11 countries. Users can use this program to independently and safely navigate through various locations such as university campuses, airports, hospitals, public institutions, and shopping malls. The system utilizes Bluetooth Low Energy (BLE) beacons, produced by Boni Company, that have a capacity of 2700mAh batteries and operate in the 2.4 GHz license-free band. These beacons have Apple MFI certificates, are compatible with Bluetooth 4.0, and can be used on iOS and Android. The system can achieve an accuracy of 3-5 meters depending on decreasing beacon powers and increasing the number of devices. In some corridors, accuracy can even be as close as 3 meters, allowing the user to be directed to the door of an office or adjacent office doors. However, additional routing systems may be necessary to enhance the indoor navigation system, which will be evaluated in more detail through user feedback.

Farooq Shaikh [4] describes a prototype system designed to alleviate the difficulties faced by the blind, which uses ultrasonic sound, a gyroscope sensor, GPS, and GSM technology, as well as an Arduino board. The system allows blind individuals to navigate unknown environments, tracks their location in real time, and updates it on a web portal for family members

to monitor. Additionally, the system includes a panic button that sends an alert message to registered family members and local police in case of an emergency.

The system comprises four main technological modules: Arduino Nano, Ultrasonic sensor, GSM, and GPS. The Arduino Nano is a microcontroller that coordinates all the modules to work together seamlessly. The Ultrasonic sensor detects the distance of an object in front of it by emitting ultrasonic sound and measuring the time it takes for the sound to reflect. The GPS module provides the GPS coordinates of the person wearing the device, which can be tracked in real-time. The GSM module allows communication between the device and registered users through sending and receiving messages with the device's location. The system uses a SIM900 GSM module and a Parallax GPS module. The proposed system offers a number of advantages for visually impaired individuals. First, The ability to navigate through familiar and unfamiliar environments using ultrasonic sensors, which can greatly improve their ability to independently navigate and reduce their dependence on others for assistance. Second, A panic button that, when pressed, sends an alert message with the user's location to registered family members and the nearest police station, providing a sense of security and the ability to quickly respond in case of an emergency or threat. Third, continuous tracking of the user's location and real-time updates to a web portal allowing family members to have knowledge about the user's location and safety, which can be particularly helpful if the visually impaired person is a child. This system can improve the quality of life for visually impaired individuals by helping them to navigate safely and providing a sense of security and peace of mind for their loved ones.

Li-fi technology is a method of transferring information using LED lights. It is a fast and cost-effective remote communication system compared to Wi-Fi. It provides high security, large amounts of data transfer, and is low cost. This technology was developed as a response to the limitations of RF data transfer. Indoor navigation is essential for everyone, but especially for the visually impaired. It utilizes an unlicensed frequency range that is not hindered by RF interference. Additionally, most indoor spaces have ample sources of sunlight and provide extra safety as Li-Fi cannot pass through objects. Monika Ramchandra Botre [5] proposed a system utilizing LED lights to produce distinctive light and location data, which is received

by an embedded device or mobile phone. The device or phone then calculates the best way to reach a destination and gives vocal directions to a blind individual.

The proposed system utilizes Li-Fi technology, which transmits information through illumination using LED lights. The system makes use of LEDs that are already present within an indoor infrastructure. A navigation structure using Li-Fi Receiver is developed. In this system, an LED light is connected to a controller that processes the information. Small changes in amplitude can manipulate the intensity of the light to convey information. The technique involves rapid ON and OFF switching of the LED light, allowing immediate data transmission. The wall or ceiling has a transmitter unit that modulates the data and sends it via the LED. The receiving device is a photo-junction transistor that receives the information from the transmitter-connected LED. The data includes location information, and an associated location message is sent to the receiver whenever the receiver module enters that transmission room. The message is then processed by a microcontroller and synthesized into an active speech to give direction to the individual and a vibrator engine to guide the person.

The system offers several advantages. First, Li-Fi uses light to transmit data, which eliminates many of the issues caused by magnetic waves outside the color spectrum and can provide a more secure and reliable communication method. Second, the proposed system includes location-based services and an internal visual information mechanism using LEDs, which are projected as a route guidance system for blind people. Third, the system consists of obstacle detection using an IR sensor and a vibrator motor to alert visually impaired people. In addition, the system also uses output in an audio format to guide individuals with visual impairment and allows for voice inputs to make navigation more effective and accurate. The proposed system can improve the quality of life for visually impaired individuals by helping them to navigate safely and providing a sense of security and peace of mind using Li-Fi technology, voice-based navigation, and obstacle detection.

Hakar Mohsin Saber [6] proposed a design for a wearable system that would assist visually impaired individuals in navigating both indoors and outdoors. The system would utilize

computer vision and deep learning to identify objects and ultrasonic sensors to detect obstacles, providing audible communication of object names and vibration alerts for nearby obstructions. It would also include GPS functionality to share the user's location with a designated caregiver. The system would be controlled by an Arduino Mega microcontroller programmed with efficient algorithms. The proposed system aims to increase visually impaired individuals' awareness, independence, and safety.

The proposed system is designed to assist visually impaired individuals in navigating and monitoring their health. It consists of two main sections: a hardware section for navigation and health monitoring and an image processing and deep learning section for identifying and reading objects. The hardware section includes an Arduino Mega 2560 microcontroller, a GPS shield, a GSM module, an ultrasonic sensor, and a vibration motor. The microcontroller acts as the "brain" of the system, gathering data from the other components, processing it, and providing output. The GPS shield receives the user's location and sends it to the microcontroller for processing. The GSM module allows communication between the user and the system, sending data to a designated caregiver when necessary. The ultrasonic sensor detects obstacles and alerts the user through the vibration motor. The image processing and deep learning section use a computer with MATLAB installed to recognize and read objects within video frames recorded by an attached webcam. The system also includes a wireless headphone that reads the object names to the user, with a battery life of around five hours and an additional charging case battery of 18 hours.

The proposed system has several advantages in terms of reliability and efficiency. The system has been tested and calibrated, ensuring that all functionalities and parts are reliable. The system is also designed to be fast, with an average time per cycle of less than 400 milliseconds. Additionally, the ultrasonic sensor is set to notify the user of obstacles at a 1-meter distance, prioritizing obstacle detection at the beginning of every cycle, which allows the user to be notified with enough space before reaching the obstacle. The object recognition feature is a key aspect of the system, as it gives visually impaired individuals more confidence by knowing the objects around them. The system uses the AlexNet deep CNN for object recognition, which has an accuracy of 85% for recognizing up to 1000 objects. However, the system's accuracy

does drop when the frame contains several objects close together or when the object is not in a good position in the frame. The system also cannot recognize objects in every environment, such as environments where there are multiple objects close and adjacent to each other, where the lighting is not sufficient, or the object is not facing the camera and is placed in an angle where it looks like a similar object.

Nitin Kumar [7] proposed a navigational system to assist visually impaired individuals in navigating both known and unknown environments. The system uses a live feed, a YOLO-based algorithm, and a stick for detecting objects to help the user navigate. The model was trained on 300 images of 25 classes, and the path detection was executed through a novel tracker system using an offline-trained neural network. The accuracy of the proposed model was found to be 81% when using a wearable mask and 96.14% when using the stick.

The proposed system includes a wearable device composed of a Raspberry Pi, a camera module, and an audio jack. The device provides feedback to the visually impaired person to help them navigate safely through any dynamic environment. The system uses the You Only Look Once (YOLO) algorithm for detecting real-time objects, specifically YOLO V3, which is an improvement over previous versions in terms of accuracy. Also, it uses a python script to access the Open VINO environment, read frames from a video stream, and perform near real-time object detection using a Raspberry Pi and YOLO. Additionally, the device uses an Ultrasonic sensor, a Pi Camera, headphones, a power supply, and a 3D-printed structure to encapsulate the hardware components. The proposed system has several advantages. First, it is trained on real-life scenarios and is able to detect and track objects for the user. Second, it is cost-effective, as it does not require using expensive LiDAR or sonar sensors. Third, it has an accuracy of 81% with the wearable mask and 96.14% with the stick, and it handles past deficiencies of low precision and delivers output in faster times. However, there are still some issues to be addressed in the future, such as incorporating Robot Operating System (ROS) with a LiDAR sensor to get a more robust architecture for the navigational task.

Saifuddin Mahmud [8] proposed the design and development of a personal assistant robot for visually impaired individuals. The robot is controlled by voice commands and uses a core-lation factor (CRF) algorithm to find and determine the relative location of objects in indoor

environments. The robot is semi-humanoid and equipped with several cameras on different parts of its body for autonomous movement, object detection, distance measurement, and motion planning. The robot keeps the user informed of its actions, making it more useful. The proposed system was tested in indoor environments, and results show that the robot performs all actions with high accuracy, making the indoor environment safer, more convenient, and more comfortable for visually impaired individuals.

The proposed assistant robot is designed to aid visually impaired people in indoor environments by finding and detecting the relative location of objects. The system is controlled by voice commands through Google Assistant and coordinated by the Robot Operating System (ROS). It consists of five main modules: the physical robot, voice control, path planning, object finding, and voice feedback. The physical robot used is the TeleBot-R2, a semi-humanoid robot with 360-degree maneuverability and multi-track functionality. The voice control module utilizes Google Assistant and Dialogflow for natural language understanding and processing user commands. The path planning module employs the ”frontier-exploration” package of ROS for autonomous exploration and mapping of the indoor environment. The object-finding module uses YoloV3, a convolutional neural network, for target and reference object detection and an OCR-based package for room identification. Once an object is found, the robot stores its location for future use and provides relative information to the user. In addition, the robot has been trained with several objects of indoor environments so that it can recognize and locate objects. It also gives the necessary relative location of the object to the user successfully. Compared to other systems, including travel aid devices like white canes and guide dogs, This system can act as a personal assistant for people with visual impairment in the indoor environment and has more capabilities.

Kabalan Chaccour [9] proposed a new approach to an ambient navigation system that helps visually impaired or blind people move independently indoors. The system uses IP cameras installed on the ceilings of rooms and a smartphone as a human-machine interface. The frames captured by the cameras are analyzed by a computer vision algorithm that detects and recognizes objects, provides guidance and detects obstacles. The system is controlled by voice

commands via a mobile application. It provides feedback through voice messages to assist visually impaired and blind people with reliable indoor navigation and obstacle avoidance.

The proposed system is a new concept in safe indoor navigation with minimal complexity. It aims to provide reliable, easy-to-use, cost-effective indoor navigation for visually impaired individuals. The system architecture is simple and consists of two main components: a smartphone and a remote processing system. The smartphone acts as the interface between the user and the remote processing system, which is based on image analytic algorithms that analyze photos taken from cameras installed in the environment to compute the user's orientation and guide them to their destination safely. The mobile application on the smartphone allows the user to give voice commands, which are converted to text and transmitted to the remote processing system through a wifi or Bluetooth connection. The remote processing system then responds with navigational directions converted back to speech for the user to hear. The mobile application also has a redundant point-to-point Bluetooth connection to ensure a reliable connection between the user and the system. The system is designed to be easy to operate, requiring only a smartphone and an indoor wifi connection, and can be fixed on the user's belt to free their hands for other tasks. The proposed system offers several advantages over other indoor navigation systems for visually impaired and blind individuals. First, It is a non-wearable assistance aid, only requiring the user to carry their smartphone. Second, the cameras are attached to the ceiling of the rooms, providing coverage of the entire surface, and the user only needs to wear a marker with an imprinted pattern to be detected. Third, the system is voice-commanded, providing GPS-like navigation that is easy to operate and does not require any extra skills. Fourth, the system provides navigation assistance, obstacle avoidance and object recognition functionalities in indoor premises. Furthermore, unlike other systems, the proposed method has a simple architecture that allows users to operate completely independently at home or at work. Federica Barontini [10] proposed a novel indoor navigation system based on wearable haptic technologies and was developed with input from visually impaired individuals. The system includes an RGB-D camera, a processing unit, and a wearable device that provides force cues for guidance in an unknown indoor environment. This research article highlights the effort to improve the autonomy and quality of life for blind individuals through the development of

technological travel aids. The system utilizes a haptic interface, specifically, a wearable device called the CUFF, to deliver navigation instructions and information to the user. This allows the user to have their hands free while still receiving important commands and instructions. The CUFF device consists of two motors that create different types of tactile sensations on the user's arm, such as tightening or loosening to convey a normal force or sliding to convey tangential force cues and directional information. The device was re-engineered to be more wearable, compact, and lightweight. In addition to the CUFF, the system also uses an RGB-D camera and a processing unit to detect obstacles and provide commands to the CUFF based on an obstacle avoidance algorithm. The processing unit is currently a light laptop but could be replaced by an ad-hoc unit in future developments.

The wearable navigation system proposed in this work has several advantages over other systems. First, it uses a user-centered approach that is informed by a preliminary investigation of the requirements of visually impaired people and tight interaction with real end-users. This allows the system to be tailored to the specific needs of the users. Second, the system uses a combination of an RGB-D camera, a processing unit, and a wearable fabric-based device to convey navigational information through normal and tangential force delivery on the user's arm. This device is a new version of the cutaneous passive haptic interface described in a previous work, which has been redesigned to be more compact and lightweight. Third, the system was validated through experiments with both blindfolded participants and blind users, in different indoor environments, which showed that it could be a viable solution to increase performance of users with regard to autonomous navigation with and without the white cane usage. Fourth, visually impaired people who performed the experiment with the CUFF only exhibited a good time for task accomplishment and a positive perception of the navigation system and haptic stimuli. Furthermore, the expert users of the white cane considered it a valid aid for training newly blind people to use the white cane.

2.4 LiDAR technology in distance measurement

The use of robots in dangerous conditions can be a solution to mitigate the risks for humans and their surroundings. Dony Hutabarat [11] presents an autonomous mobile robot that utilizes a Light Detection and Ranging (LiDAR) sensor to detect and avoid obstacles. LiDAR sensor works by emitting a laser beam and measuring the time it takes for the Light to return after hitting an object, determining the distance of the object. This sensor is considered a more advanced sensor than conventional sensors as it can provide more accurate and reliable information about the environment. The robot's navigation is guided by the Braitenberg vehicle strategy, a method of creating simple but intelligent robots with minimal sensors and simple rule-based behaviors. The sensor data collection and control algorithm is implemented on a Raspberry Pi 3 computer board which allows for efficient processing of the sensor data. The results of the experiments indicate that the LiDAR sensor is able to measure distance consistently, and it is not affected by the object's color or ambient light intensity. This allows the robot to navigate effectively and avoid obstacles of different sizes and colors. Furthermore, the mobile robot is able to move around a room without any impact on the walls or other obstacles. The system uses the LiDAR sensor to obtain the coordinates of the angle and distance of the object. The sensor scans the environment in a clockwise direction and only uses 360 data in the range from 90° to -90° for a half-degree change. It is built on top of the Robot Operating System (ROS) and YDLiDAR drivers, which are installed on a single Raspberry Pi 3 computer board. In addition, it utilizes the Braitenberg vehicle 2b method, where the sensor on the left side of the mobile robot is used to control the right motor and vice versa. The motor speed is determined by the Pulse Width Modulation (PWM) method. If the distance is less than 0.5 m at an angle between 90· and 0·, then the right motor speed will gradually decrease until the distance is 0.15 m. The motor speed reaches a minimum and makes the robot turn right, and vice versa.

LiDAR is an optical scanning technology that measures the properties of radiated light to find distance and other information from a target. The system is equipped with a LiDAR sensor capable of measuring distances between 0.12 to 10.5m with an error rate of 0.9%. This

LiDAR sensor can rotate to provide a 360-degree view, and the interface data communication uses a serial port or USB adapter. The advantage of this system is that LiDAR has a high level of precision with a long detection distance, and the sensor is not affected by the color of the object or the intensity of ambient light. The autonomous mobile robot can avoid colored objects of different sizes and navigate the indoor room with or without obstacles without any impact on the wall or obstacle. Additionally, the system can be implemented on a single-board computer of Raspberry Pi 3. Thus, the use of LiDAR sensors in this autonomous mobile robot improves its ability to detect and avoid obstacles and navigate through dangerous environments, providing a safer solution for humans and the environment.

King [12] discusses how snow is essential for the global water-energy budget and how laser altimetry (LiDAR) is a useful technique for monitoring snow depth. However, traditional LiDAR equipment is expensive and difficult to transport. The authors of the study demonstrate that the LiDAR sensor on the Apple iPhone 12 Pro smartphone can be used as a real-time, handheld measurement instrument for observing changes in snow depth. Two field experiments in Canada found that the iPhone LiDAR was able to accurately capture daily changes in snow depth compared to measurements taken with a ruler. The high accuracy of the LiDAR sensor suggests that it could be used to develop a mobile application for measuring changes in snow depth through a citizen science-based approach.

The system combines in situ manual depth measurements and LiDAR sensor technology. In situ measurements are made using a 1-meter-long steel ruler inserted vertically through the snowpack to the snow-ground interface. The LiDAR sensor is built-in the iPhone 12 Pro camera module, which operates near-infrared (NIR) and uses photon-counting detectors (also known as single-photon avalanche photodiodes or SPADs) in a direct time-of-flight measurement approach. The sensor has a range of approximately 5 meters, a 90-degree field of view, and measurement accuracy within 5 mm of the true distance from the sensor. The sensor point density follows a linear trend along a logarithmic scale to produce a large number of depth points per unit area. The raw depth data is made available from the sensor and is sufficient to provide a detailed representation of the surrounding scene. The sensor operates in real-time while scanning the scene using a combination of the phone's GPS, accelerometer, and LiDAR

sensor to provide a detailed 3-D colored mesh composed of tens of thousands of data points. The advantage of using a LiDAR sensor embedded in a smartphone, such as the iPhone's iLiDAR, for measuring snow depth over other methods is that it is orders of magnitude less costly than specialized survey-grade equipment, and its portable nature enables measurements to be made in a more accessible manner. Additionally, by connecting a smartphone application with a wider citizen-science cluster network of atmospheric observations, it has the potential to improve global estimates of snow depth.

The traditional method of topographic surveying in earth sciences is costly and complex, involving significant financial investments, detailed logistics, and specialized training. However, recent advancements in technology have allowed for the use of off-the-shelf drones equipped with optical sensors to obtain high-resolution datasets of Earth surfaces at a lower cost. In 2020, Apple Inc. released the iPad Pro 2020 and the iPhone 12 Pro with built-in LiDAR sensors, which have the potential to further reduce the costs and complexity associated with topographic surveying. Luetzenburg [13] in this study investigates the technical capabilities of the LiDAR sensors and test their application at a coastal cliff in Denmark. The results show that the LiDAR sensors can accurately create high-resolution models of small objects with a side length greater than 10 cm, with an absolute accuracy of ±1 cm. Additionally, 3D models of larger areas, such as a coastal cliff measuring up to 130 x 15 x 10 m, can be compiled with an absolute accuracy of ±10 cm. The study concludes that the versatility and ease of use of the Apple LiDAR devices make them a cost-effective alternative to established techniques in remote sensing, with potential applications in a wide range of geoscientific fields and teaching.

The advantage of this system, which uses LiDAR scanners available on consumer-grade devices like the iPad Pro and iPhone, is that it is low-cost and accessible to a wider range of people, including citizens, for scientific mapping and observations. This allows for high temporal and spatial resolution and data acquisition that is not restricted by weather conditions or access to rough terrain. Furthermore, the use of smartphones allows for crowd-sourced observations and participation of citizens in science. The LiDAR sensor also has a high detection rate of tree stems above a threshold of 10 cm diameter and can be used in geoscientific research areas like coastal cliff erosion.

Keeffe [14] focuses on applying the technology used in self-driving vehicles to enhance navigation for visually impaired and blind individuals. The study aims to integrate obstacle detection sensors into an intelligent white cane, including a long-range LiDAR (up to 10m) and other main sensors. The goal is to create a cane that can scan its surroundings and provide additional environmental information to aid navigation. The authors examine the challenges and benefits of using autonomous vehicle technology in this context, including the need to reduce the weight and size of the system while accommodating slower speeds for pedestrians. This work is part of the INSPEX H2020 project.

The INSPEX system utilizes various sensors to detect obstacles in the user's path, such as Ultra WideBand RADAR, long and short-range LiDAR, ultrasound, and range sensors. These sensors complement each other to provide a comprehensive set of data for the surroundings at distances of 0-10m. However, no such integrated system currently exists due to the size and power limitations of individual sensors and challenges in multiple sensor integration. The system must be lightweight, compact, and have a long battery life of 10 hours, with power consumption under 500mW. The information on obstacle location will be conveyed through a 3D Audio interface. The 25m long-range LiDAR is a prototype developed by SensL and consists of a laser diode, time-to-digital converter, and FPGA components. It operates by emitting a high-power light pulse, measuring the reflection interval, and determining the obstacle distance. The current prototype is large due to testing elements that will not be included in future versions. The focus of the paper is on the long-range LiDAR sensor, which can detect obstacles up to 10m. The system uses a laser diode to emit a pulse of light, and a sensor detects the reflected light to determine the distance of the obstacle. The system has been tested using indoor and outdoor obstacles and shows promising results for obstacle detection up to 5m. However, in outdoor conditions with high brightness, the detection angle needs to be reduced for reliable detection up to 10m. The system will also need to address the size and weight of the device for use on a cane. The next generation is expected to have a smaller footprint and improved optics for outdoor use. The system will be integrated with short-range LiDAR, ultra-wideband radar, and ultrasound range sensors as part of the INSPEX system. Each sensor will meet power, size, and weight requirements for detection range under different environmental conditions.

P.Chitra [15] presents a new solution to help visually impaired people navigate their surroundings with greater ease. The authors propose a wearable device that utilizes LiDAR sensors and vibrotactile units to help detect obstacles and provide feedback to the user. The device uses a combination of LiDAR, a convolutional neural network, and audio output to help the user distinguish between free space and obstacles. The LiDAR sensor measures the distance between the user and obstacles and provides feedback to the user through voice input and vibrations from a vibratory motor. The goal of the device is to provide a safer, more comfortable alternative to traditional white canes for the visually impaired.

The system is designed using Raspberry Pi, LiDAR, Camera, a Motor, a Relay module, a wearable strap, Headphone, a USB cable, and a Power supply unit. The Raspberry Pi 3 is used as a processor to read the distance data from the LiDAR sensor. LiDAR is used for remote sensing to calculate the distance of objects. The vibratory motor is used to vibrate when the obstacle is detected. The camera is used for streaming video and capturing images for image processing. The haptic strap holds all the hardware components, and the audio jockey informs the name of the obstacle detected to the blind through audio. The proposed system provides a more comprehensive solution for safe navigation for visually impaired people compared to other existing systems. It uses a combination of LiDAR, vibrotactile motor, and computer vision technology for high accuracy and provides information about the obstacles in terms of distance and object details. The wearable device with sensors and haptic devices is lightweight and easy to use, making it suitable for day-to-day use. The system uses the Raspberry Pi as the central component, allowing for easy connections to other devices and monitoring of the user's haptic strap. This system offers a more advanced solution compared to systems relying solely on ultrasonic sensors or GPS technology and has been tested and refined to provide practical and efficient assistance to visually impaired people.

The eyeDog assistive-guide robot [16] is designed to provide visually impaired people with autonomous vision-based navigation and laser-based obstacle avoidance capabilities. The system consists of the iRobot Create platform, a notebook computer, a Logitech USB webcam, and a Hokuyo LiDAR unit. The camera serves as the primary exteroceptive sensor and is used to estimate the position of the vanishing point from the captured video, which is processed

using OpenCV. The deviation from the principal point of the camera is calculated, and the robot steers accordingly to move parallel to the direction of the road. LiDAR is used for obstacle detection and avoidance. The software architecture is distributed at the component level and is networked, allowing for easy integration of new software modules. The system is cost-effective and easily built, making it a promising solution for the visually impaired. Future developments for the eyeDog include a more rugged and robust mechanical platform, a PID controller for translational velocity, and a user interface specifically tailored for the visually impaired.

The proposed system offers several advantages over existing assistive devices and traditional guide dogs. First, The average cost of a guide dog is $42,000, including training, whereas the proposed system is expected to be more affordable. Second, the robot can perform more complex commands and remove the burden of high-level Navigation from the user. Third, the system uses computer vision techniques and laser data to navigate indoors and outdoors, whereas previous attempts have only focused on indoor Navigation. Fourth, unlike previous systems, the proposed system does not require any special modifications in existing stores. Fifth, the system uses RANSAC with adaptive thresholds for robust clustering of multiple lines, making it suitable for real-time Navigation. In addition, No need for training: Guide dogs require extensive training for both the dog and the user, whereas the proposed system does not require any training.

Zhang [17] proposed a multi-task perception system for scene parsing and recognition for individuals with visual impairments. The system is built on a compact ResNet backbone and features two paths with shared parameters for improved efficiency. One path focuses on semantic segmentation, utilizing fast attention to gather contextual information, while the other path performs scene recognition through a combination of semantic-driven attention and RGB representations. The system was tested on both public datasets and real-world scenes, showing good accuracy and efficiency. The system is designed to be wearable and can provide assistive scene information to aid in navigation tasks through the use of an Intel RealSense LiDAR camera and an Nvidia Jetson AGX Xavier processor.

Ton [18] presents a novel solution for visually impaired individuals to comprehend their surroundings through sound. The traditional method of echolocation requires extensive training and can be impacted by various conditions. The proposed solution, LiDAR assist spatial sensing (LASS) system, utilizes a LiDAR sensor to gather information about the environment and translate it into a stereo sound of different pitches. This sound provides information on the location and distance of obstacles, offering visually impaired users improved spatial awareness. The efficacy of the system was tested with 18 student volunteers, and results showed that with minimal training, they were able to accurately identify and distinguish multiple objects. The study was approved by the Penn State Institutional Review Board.

The LASS (LiDAR Assisted Spatial Sound) system employs a Hokuyo URG-04LX Scanning Laser Rangefinder as the main detection device to collect spatial information in a semi-circle area in front of the user. The LiDAR scans between 0-180 degrees, emitting an infrared laser beam of 785 nm wavelength, which is classified as safe (Class 1 laser). The data collected is delivered to the user through stereo sound headphones, giving directional information about obstacles. The system maps distances to audio frequencies, with closer obstacles having a higher pitch. The audio frequency range is set to differentiate spatial information easily for the user. The system operates every 13 seconds and processes the data gathered. Compared to other systems, the LASS system utilizes LiDAR, a device with a shorter wavelength and focused beam, resulting in higher spatial resolution compared to ultrasound. It does not require head movements from the user for optimal performance, as LiDAR constantly scans the area in front of the user. Also, it translates distance information into frequency-related sound signals that can be optimized to the user's reaction, making it easier for users to gauge distance. It is mounted on a chest harness that turns with the user, allowing the LiDAR to collect spatial information right in front of the user constantly. In addition, it requires less training time and can be dynamically optimized for each user, and generates and receives signals, allowing the user to focus on interpreting the spatial information.

Busaeed [19] presents a proof of concept for a new wearable technology, LidSonic, aimed to assist visually impaired individuals with navigation. The smart glasses use machine learning, LiDAR, and ultrasonic sensors to identify obstacles in real-time. The system, consisting of an

Arduino Uno device and a smartphone app, detects objects and provides buzzer warnings or verbal feedback to the user. The design focuses on affordability, reliability, low energy consumption, and simplicity. The evaluation of the system using nine machine learning algorithms showed promising results, and the complete system was built using inexpensive off-the-shelf hardware and software tools. This work provides a valuable contribution to the field of assistive technologies for the visually impaired and opens new directions for the design of smart glasses.

The LidSonic system is a hardware and software solution for visually impaired people. It consists of a device mounted on glasses containing HC-SR04 ultrasonic sensor, TFmini-s LiDAR, laser, servo, and Bluetooth connected to an Arduino Uno board. The device serves as the system's "brain", integrating and managing the sensors and actuators and communicating data to the smartphone app through Bluetooth. The ultrasonic sensor detects obstacles within 30 degrees and 0.02-4 m range, and the TFmini-s LiDAR collects data from its environment. The smartphone app has a microphone for voice commands, Bluetooth for communication, and speakers for verbal feedback. The software is composed of four modules: Sensors, Dataset, Machine Learning, and Voice. The Sensors module manages the sensors and actuators, while the Dataset module collects and stores data. The Machine Learning module provides model training, inference, and evaluation. The Voice module uses Google Text-To-Speech and Speech-to-Text APIs to provide audio feedback and voice commands. LidSonic system uses machine learning, LiDAR, and ultrasonic sensors to identify obstacles. It is affordable and inexpensive at the cost of less than USD 80, which is Lightweight and easy to use. It uses simple data processing for faster inference and decision-making and can be built into or mounted on any pair of glasses or wheelchair. In addition, it improves the quality of life for visually impaired people and can be enhanced with more data for classifier training. However, it is currently only a prototype with room for performance improvement, and only four people tested the system, further testing with visually impaired users is needed. The voice feedback system needs further study to provide convenient notifications.

The recent advancements in the robotics industry have had a significant impact on the visual path-guiding sector. Many researchers have proposed various blind assistive systems.

However, guiding a mobile guiding robot through an unpredictable environment is still a challenge. IoT-enabled robot control offers a more flexible and effective form of communication in the robotics industry. The system created by Kalpana [20] is a wireless smart robot that is meant to assist visually impaired individuals in navigating indoors. This robotic dog can guide visually impaired people to a locally programmed destination by recognizing words spoken by the user using Google Voice Recognition API. The robot also has a light detection system to detect obstacles in its path and uses a Corner Crossing Algorithm to avoid corners in its operating environment intelligently. The robot also has a ”watchdog” mode, which acts as a watchdog and warns the user of any abnormal movement, and is equipped with a fire sensor. Additionally, the robot is designed with a self-charging feature that allows it to find sunlight and charge itself during the day using photovoltaic cells.

The robot uses a combination of sound navigation and passive infrared sensors to avoid obstacles and a virtual map of the home to guide the user to their desired destination. The heart of the system is a microprocessor that uses RTOS for efficient processing. The user communicates with the robot through a guide stick and Bluetooth 4.0 connection. The app converts voice commands into text and sends them to the robot. This system provides a Voice recognition feature that allows visually impaired individuals to control the robot easily. It utilized Bluetooth 4.0 connection between the robot and Android interface to improve communication and advanced geographical sensing model to select a path and avoid obstacles. In addition, the Android user interface is designed to be user-friendly for those with limited technical knowledge. Moreover, the Self-charging feature eliminates maintenance for visually impaired individuals. However, the average working lifespan of the robot is limited to 5-6 years. And temporary inbuilt battery has limited operating power. It may have difficulty recognizing voice commands or interpreting them accurately. In addition, reliance on sunlight for charging may not be ideal in all locations or weather conditions.

Dragne [21] presents research results of an evaluation of object detection and recognition techniques specifically aimed at developing assistive devices for visually impaired persons. The study focuses on two techniques: one using low-cost photo cameras and sign detection, and the other using a LiDAR sensor and an HQ photo camera. The results of the experiments showed

the superiority of the latter technique, with high precision and fast recognition of objects at short distances. This research presents two novel aspects. Firstly, it introduces a method for distance assessment using low-cost cameras based on special sign detection. Secondly, it presents the development of a tailored perception subsystem using a laser RPLIDAR sensor and a photo camera. A new technique for distance assessment is presented that utilizes specially designed signs. These signs are marked with distinctive geometric shapes, colors, and shading to make them easily recognizable. The distance to objects is determined by recognizing these signs with a camera and calculating the disparity between the captured image and reference images stored in a database. People with visual impairments are referred to as ”Explorers”. The signs are designed to be rich in color and simple in shape for easy recognition and a positive effect on Explorers.

2.5 Accessible User Interface Design

In recent years, mobile devices have become essential for daily communication, work, and entertainment. At the same time, the need for accessibility of technology has escalated. According to the World Health Organization(WHO), as the population ages and chronic health conditions increase, the number of people with disabilities will increase. However, the key to solving the disability problem is not to cure the ”lack” of disabled people's limbs, vision, hearing, and other functions in a medical sense but to integrate the differences of disabled people into social norms based on respect and rebuild an open, integrated and friendly social environment for members of this group. In this background, the concept of accessibility came into being and became the consensus of the international community. Countries have already provided institutional guarantees for accessibility through legislation. The world's first reference style book on accessible environment design, ”Sources of Ideas for Accessible Design,” was developed by the United States. Accessibility is defined as ”the extent to which groups of people can use products, services, environments, and facilities with different characteristics and abilities to achieve specific goals in specific usage contexts.” Accessibility includes human environment accessibility, material environment accessibility, and information and communication accessibility. Whether users have a permanent, situational or temporary disability, accessible design

Figure 2.1: Sidewalk ramp accessibility.

means being inclusive of all. A classic concept is the sidewalk ramp phenomenon. It refers to something designed for people with disabilities, often to help everyone. For example, a sidewalk slope refers to a hill where the sidewalk meets the road, see Figure 2.1.

You must have seen sidewalk slopes, but you probably never realized they existed. In most countries, sidewalk ramps are required by law to help people in wheelchairs get around. But in reality, it can also help many other people, such as travelers with suitcases, people pushing prams or bicycles, people on skateboards/scooters, couriers pushing trolleys, etc. If you've ever traveled with luggage, you must benefit from this accessible design. Other examples of the sidewalk ramp phenomenon include: (i) Email: Vint Cerf created the first commercial email system in the earliest days of Internet development. His motivation for designing email was that Vint Cerf was hearing impaired and relied heavily on written communication. Today, we will use email more frequently for our daily correspondence. (ii) Typewriter: The earliest typewriter was invented by Italian inventor Pellegrino Turri in order to allow his gradually blind lover to write neatly and beautifully when writing love letters. [43] As we know, the typewriter has become the keyboard we use every day. (iii) Subtitles were first created to help deaf and hard-of-hearing people watch TV programs. However, many people with a good hearing now also turn it on when watching videos. And since there are many Asian countries with many different dialects spoken in the region, you may find that many TV shows in these countries

have embedded subtitles. The subtitles allow people of different dialects to understand the content of TV programs. In addition, accessibility is a magnifying glass that helps you see potential problems in your product. If someone with tremors has difficulty reaching your button, it may be easy for many users to touch it by mistake. If a person with a learning disability says your interface and layout are challenging to understand, then the interface is likely to be a bit complicated for many users.

Visual impairment, also known as impaired vision and loss of vision, refers to the decline in vision to a certain extent, which cannot be corrected by general methods such as glasses, and also includes those whose visual ability is impaired because they cannot wear or own glasses or contact lenses. Visual impairment can cause people to have difficulty with activities of daily living, such as driving, reading, socializing, and walking. According to the degree of disability, it is divided into two categories: blind and low vision. Low vision refers to the best-corrected visual acuity of the eye with a measured value of more than 0.05 and less than 0.3. People with low vision are not completely lost because of their visual function if they receive appropriate assistance, such as using large-font textbooks or visual aids. Blindness refers to those whose best-corrected visual acuity is less than 0.05. Because the blind cannot perceive the external world through the visual senses, they must experience it through other senses. After entering the era of smart mobile devices, many developers helped the VIB group to design some accessibility functions to help them use smartphones. Such as blind users can read and write operations through voice commands, use cameras for banknote recognition, or use map applications to navigate. Under the blind assistance function of the iOS operating system, blind people can even recognize the surrounding scene through the camera and experience the fun of taking pictures in person through voice prompts. In daily life, blind people use common apps like healthy people, such as browsing news and listening to songs online. They also use Facebook Massage, Whatsapp, Twitter, and other social communication software to communicate with their family and friends. Since the 1970s, the computer graphics interface has begun studying human-computer interaction for the visually impaired and the blind. To facilitate the use of computers by visually impaired people, software engineers have gradually started to develop accessibility aids for the blind.

Currently, the research on accessibility assistance for visually impaired groups is divided into three parts. The first category is the ”read screen software” type of accessibility assistance functions that assist in reading screen information, the second category is computer software designed for blind users, and the third category is the study of ”blind input.” At present, we have entered the era of mobile Internet, and smart mobile devices have become the primary trend of future mobile communication development. Applications based on smart mobile devices are constantly affecting all aspects of people's clothing, food, housing, and transportation. Therefore, accessibility assistive functions for smart mobile devices have also been developed to help blind users operate and use them. In April 2009, Apple introduced a ”Voiceover” voice assistant program mainly used by various iPod players. Voiceover enables the user to long press the HOME button while playing music with the iPod so that the music player can use voice to prompt the user of the music being played and the performer's name while playing the music. In addition, music playlists can also be played accordingly. It makes playback information more accessible to the visually impaired. At the same time, by opening the ”Voiceover” function, the user can touch the text area to make the program read the text on the screen aloud and double-click to enter the text connection. Accessibility features for the intellectually disabled and blind, such as screen readers, multimodal interactions, somatosensory vibrations, haptic feedback, and gestures, are helping blind people operate touchscreen interfaces on mobile devices. Some of the usability issues with today's touchscreen user interfaces result in trade-offs in discoverability, navigation complexity, cognitive overload, layout persistence, cumbersome input mechanisms, accessibility, and cross-device interaction. One solution to these problems is to design an accessible blind user interface for everyday activities on smart mobile devices. In the research on the design of accessible user interfaces for the visually impaired and the blind, we found that at the beginning of the invention, we should consider the following aspects of the plan:

2.5.1 Text Contrast and Image Contrast

Figure 2.2 (right) shows that the color of words or icons is too light, which can cause low-vision users to see clearly. The specified in the WCAG (Web Content Accessibility Guidelines)

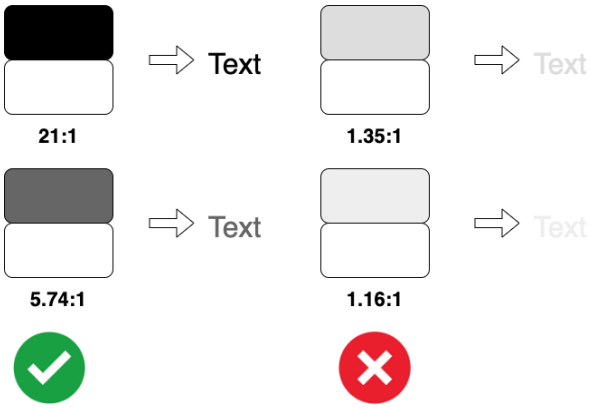

Figure 2.2: Text Contrast in WCAG.

and Apple/Google's accessibility guidelines: the color contrast ratio of all important content (including text and icons) needs to be 4.5:1 or above (When the font size is greater than 18 the color contrast ratio of the content needs to be 3:1 or above). Figure 2.2 (left) and Figure 2.3 shows the correct sample.

2.5.2 Touch Target Size

If the buttons in the interface are too small, they are not easy to click, which can make it extremely difficult for many people with disabilities to use. For example, users with physical disabilities cannot click the small button due to hand tremors. And the VIB users cannot find and accurately click the small button due to poor vision. Therefore, the size of the Touch Target is standardized as follows: Web page: clickable The area cannot be smaller than 44x44 px [40] iOS: cannot be smaller than 44x44 pt [41] Android: cannot be smaller than 48x48 dp [42]. Figure 2.4 shows the correct example.

2.5.3 Accessibility Label

Many visually impaired users must use screen-reading software when using mobile phones or computers. Screen-reading software reads everything on the page aloud, so users can "listen" to know what's on the screen without seeing the screen. Screen-reading software can directly read the text in the interface, but the picture cannot be read. The developer of the app or website needs to manually add an accessibility label (also called alt text) so that the user can know what

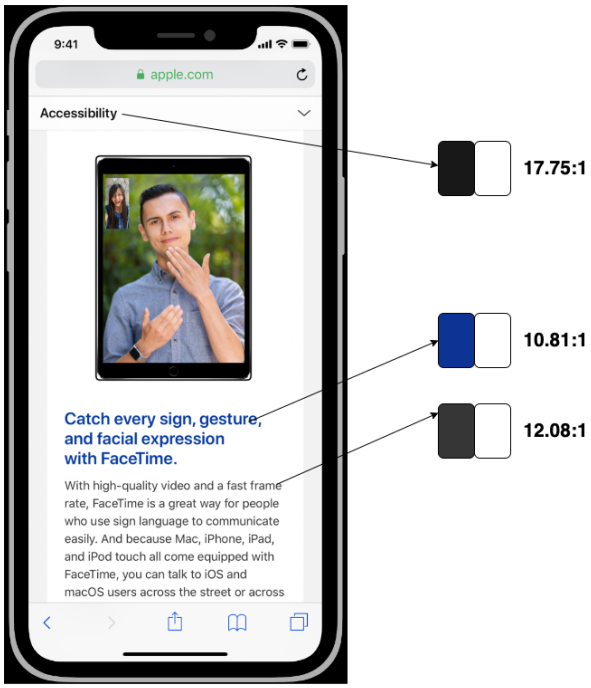

Figure 2.3: Text display from Apple's accessibility.

Figure 2.4: Touch targets in Apple: On touch screens, provide ample touch targets for interactive components. Maintain a minimum tappable area of 44x44 points for all controls.

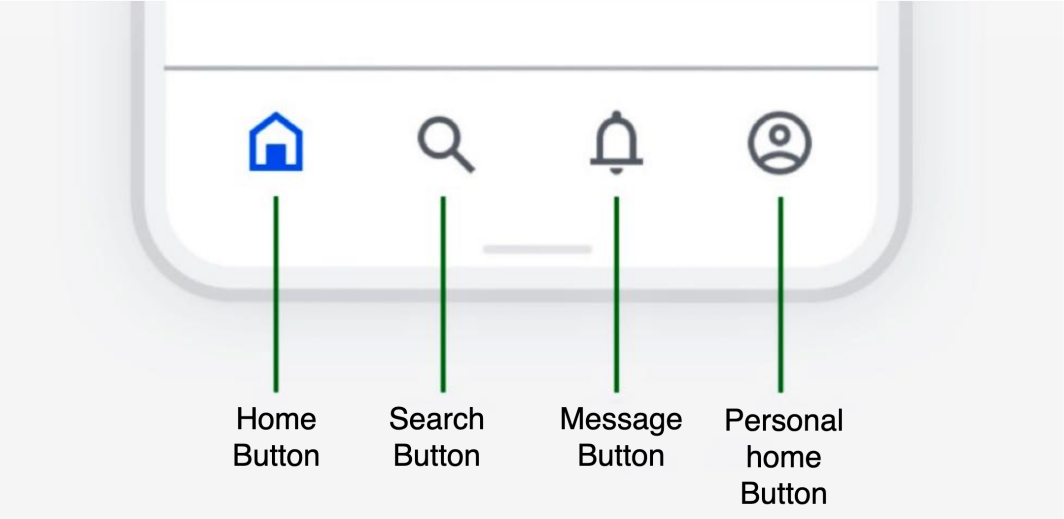

Figure 2.5: Accessibility Label.

it is. For example, the common app's bottom navigation bar has four icon buttons, such as in Figure 2.5. The user with normal vision might be able to know what each button is. But if the accessibility label is not added, VIB users using screen reading software will hear "button, button, button, button" when listening to this row of buttons. It is entirely useless to use this product. In contrast, VIB users using screen readers will hear the "home button, search button, message button, personal home button" after adding the accessibility label. It helps VIB users to select the correct button and go to the desired page.

2.5.4 Voice reminder and somatosensory vibration warning

For the visually impaired and blind (VIB), hearing and touch are essential senses for them to perceive changes in their surroundings. VIB people can quickly get the information they need through voice and make corresponding responses. Somatosensory vibration can make VIB people feel unaffected directly in a noisy environment and receive obstacle avoidance warnings and other important feedback information. These ensure their safety.

Chapter 3

NAAD system

3.1 Introduction

Visually impaired individuals face difficulties in accurately perceiving their surroundings due to their eyesight limitations, which increases their risk of indoor and outdoor accidents. To address this issue, we propose a mobile system called NAAD. It utilizes a Voice Interaction system combined with LiDAR technology, Deep Learning technology, and computer vision techniques such as Mobile-Net Single-shot Detection (MobileNet-SSD), Augmented Reality (AR) to detect objects and calculate distances. The NAAD system serves several purposes: (i) it can quickly identify user-specified items, (ii) it aids visually impaired individuals in navigating their daily lives by helping them avoid obstacles, and (iii) it integrates LiDAR for both AR and VR experiences, reducing the need for additional equipment and improving the accuracy of obstacle detection. By improving their ability to identify obstacles in their environment, the NAAD system can significantly enhance the quality of life for visually impaired people. Experimental results have shown that the NAAD system achieves a distance accuracy of 96% within a range of five meters, outpacing other research [23]. Additionally, the system operates at a rate of over forty-four frames per second, surpassing similar projects [34]. These further demonstrate its effectiveness and potential as an assistive technology for the visually impaired.

3.2 Problem overview

A report by the World Health Organization in 2019 revealed that over 2.2 billion people worldwide are affected by visual impairment or blindness. Researchers have explored various navigation tools to assist the visually impaired, including canes, navigation dogs, depth cameras, and radio-frequency, ultrasonic, and infrared sensors. However, most of these tools have limited effectiveness in indoor settings and close proximity situations.

AI-Guide [45] is a widely used solution based on the Apple SDK (ARKit) that enables visually impaired individuals to locate and grasp objects. However, this system cannot accurately identify common everyday items such as shoes, apples, and bottles. To address this issue, the research project will improve the system's small object recognition capabilities. The project will incorporate a machine-learning model that can recognize small 3D objects effectively. This approach is expected to enhance the system's ability to identify and locate smaller objects that were previously difficult to detect, thereby improving its overall usefulness and effectiveness as an assistive technology for visually impaired individuals. We also researched Microsoft's Seeing-AI, which utilizes Augmented Reality and LiDAR technology to assist with identifying text, color, currency, and other visual information. This research project will improve its functionality to support the real-time detection of general 3D objects. To enable safe navigation for visually impaired individuals, researchers have employed neural network technology, such as Convolutional Neural Networks (CNNs), to identify obstacles in their surroundings. In one study, Faster-R-CNN was incorporated into the image recognition module of a server, while the network protocol used was RTP in fast mode. Yolo was used as the image recognition module, and the Mono-SLAM formula was utilized in the direction and distance module to calculate the distance between the camera and the target. This approach was documented in [24].In another study, two neural networks were utilized: the Vision Disparity Network and the Semantic Segmentation Network. The former generated a dissimilarity map to calculate the distance between the camera and the target, while the latter was used to detect obstacles. This research is discussed in [23] [34] used the SSDLIFE MobilenetV2 model to identify obstacles, while Stereo Vision was utilized to estimate the actual distance using camera parameters and

disparity. The author utilized a Fuzzy Algorithm to classify the obstacle and object and communicate the information to the user via the Text to Speech engine. The experiment shows that high frame rates lead to increased computation. It prompted the researchers to optimize the frame rate to 15 FPS.

In the study conducted by Jiang [27], an obstacle detection method was introduced, utilizing the ALEXNET neural network. This approach successfully addresses the challenge of image recognition affected by variations in lighting conditions.

In this research, we discussed the design of a mobile device that utilizes a virtual assistant to interact with users. This virtual assistant offers various modules, including object detection, obstacle detection, distance estimation, and an Alert System that analyzes the real-time environment and alerts users to potential obstacles. To improve accuracy, we incorporate the LiDAR sensor, which outperforms the ARKIT API. Our experiments demonstrate that our system can estimate distances with 96% accuracy within a five-meter range, surpassing [23] [33] [28]. Additionally, our system's FPS reaches over forty-four frames per second, which outperforms similar projects [34].

3.3 Query mode in NAAD system

3.3.1 Research Problem

The most common challenge for visually impaired and blind (VIB) people is finding daily necessities or misplaced personal items when interacting with their surroundings.

3.3.2 Research Questions

- How to search for user-specified items within a certain area?
- How to enable our system to recognize more common objects than existing systems?
- How to design the system so users can interact with the device accessibly?

3.3.3 Research Hypothesis

- The NAAD system can help the visually impaired and blind (VIB) find essentials or misplaced personal items quickly.
- The Query Mode of the NAAD system can quickly find the object item that users need to query and notify the location and distance of the object item in real time.
- Our NAAD system can identify more common objects than existing systems.
- Users can accessibly interact with the device while using the NAAD system.

3.3.4 Object detection

3.3.4.1 Instruction

Object Detection is a significant and fundamental branch of computer vision. Object detection has been applied to all aspects of people's daily life, such as face recognition, intelligent video analysis, vehicle counting, retrograde detection, license plate detection and recognition, automatic classification of photo albums, etc. British neuroscientist and physiologist David Courtenay Marr, the founder of computer vision theory, believes that the problem to be solved by computer vision is"What is Where." It is similar to human recognition of image content. Its main task is to finally complete the image's classification, image positioning, target recognition, image segmentation, etc., through the statistics of pixel distribution, color, texture, and other characteristics.

Figure 3.1 very clearly describes several computer vision tasks and the relationship between them, namely: (1) Classification: The input is a painting, and the classifier needs to give the category assignment of the subject described in this picture. For example, the "protagonist" in Figure 3.1 is a cat. (2) Localization: Classification can only tell us the category of the picture, but it does not point out the specific position of the subject in the picture, which is done by Localization. Usually, we will study classification and positioning as an overall algorithm to directly output the complete information, such as the subject category and its position and size in the picture. (3) Object Detection: The previous classification and positioning are aimed

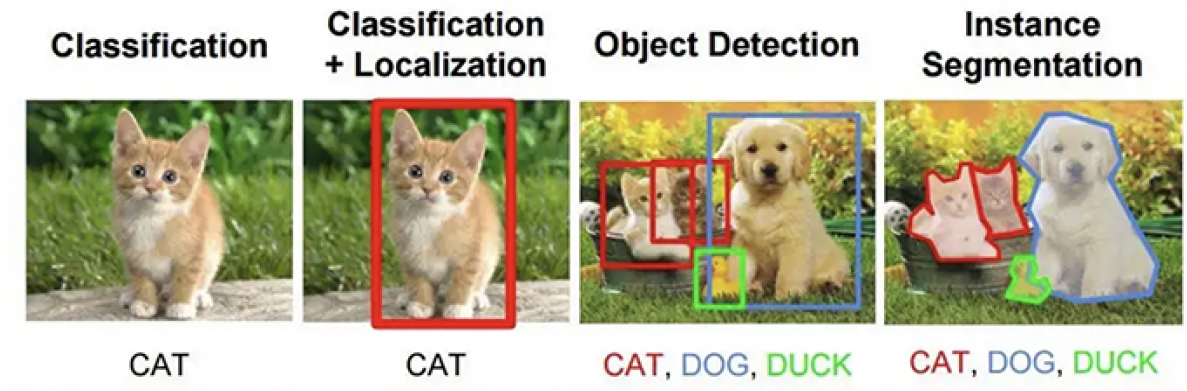

Figure 3.1: Computer Vision Tasks.

at the situation where only one subject is in the picture, which is not enough. For example, when multiple animals such as cats, dogs, and chickens appear in the image simultaneously, it becomes a problem of ”multi-target” recognition. What we call target recognition is multi-object + classification + positioning. It is also the most widely used visual recognition scene in daily life. (4) Instance Segmentation: The Instance segmentation algorithm needs to know which objects are in a picture and their positions and accurately define the objects' outlines.

3.3.4.2 Problem overview

Object detection is a very challenging problem in vision research. The difficulty of object detection is mainly divided into three levels: instance level, category level, and semantic level, as shown in Figure 3.2.

For the Instance level, a single target object instance, usually due to the different lighting conditions, shooting angles, and distances during image acquisition, the non-rigid deformation of the object itself, and the partial occlusion of other things, the apparent characteristics of the object instance are very different. Significant changes have brought great difficulties to the visual recognition algorithm.

To the Category hierarchy,First, there are significant differences in the apparent features between target objects of the same category.

For example, as shown in Figure 3.3, the same chairs have different appearances. From the perspective of people's cognition, all appliances with the function of ”sitting” can be called

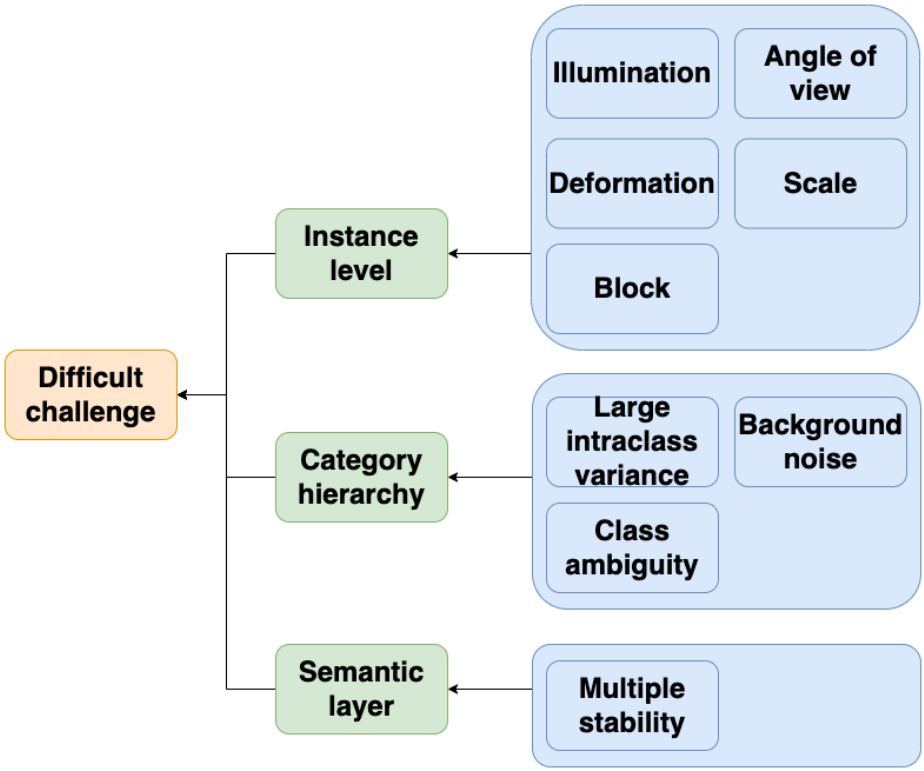

Figure 3.2: Difficulties and challenges in object classification and detection research

Figure 3.3: Significant differences in the apparent features between target objects of the same category (source: https://www.schooloutfitters.com/)

Figure 3.4: Significant differences in the apparent features between target objects of the same category (source: https://www.loveyourdog.com)

chairs. Secondly, item categories are ambiguous. The different types of objects will have certain similarities. As shown in Figure 3.4, the one on the right is a wolf, and the one on the left is a husky, but it is difficult for us to separate them from the appearance; again, the background interference, the background of the object in the actual scene is complex. It will seriously interfere with target recognition and make target recognition more difficult.

For the Semantic layer, the difficulty at the semantic level is a great challenge, especially for the current state of the art in computer vision theory.

A typical problem is called multiple stability. As shown, the left side of Figure 3.5 can see two people facing each other or a candlestick; the right side can be interpreted as a rabbit or a duckling. There are different interpretations of the same image. It is not only related to physical conditions such as people's observation angle and focus but also people's personalities and experiences. It is the hardest part for visual identity systems to handle.

3.3.4.3 Related Study

The image recognition algorithm is a significant and essential branch of computer vision. In deep learning, the image recognition model not only completes its tasks but also serves as a feature extraction network for other computer vision tasks.

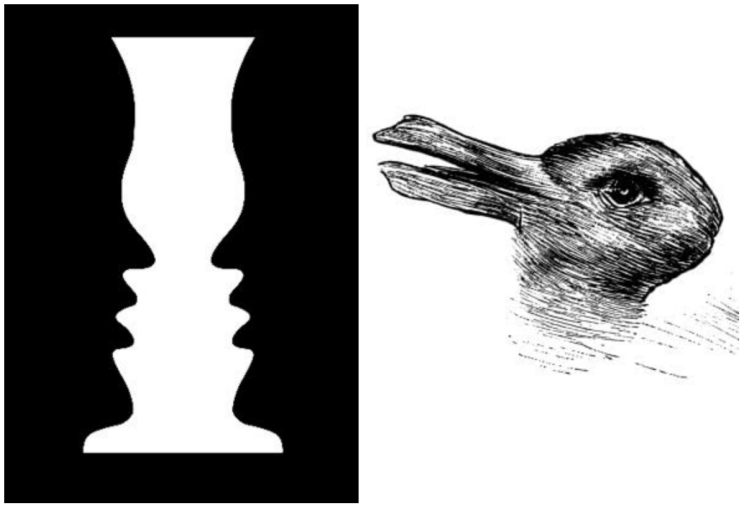

Figure 3.5: Semantic layer Difficulty at the semantic level .

3.3.4.4 Traditional Algorithm Ideas

The traditional algorithm is an algorithm that is closer to human understanding and cognition of things. For example, when teaching children to recognize objects in images, we first summarize the target objects to be distinguished as the critical basis for judging the categories. We call such judgments based on ”features.” When we want to differentiate between triangles and five-pointed stars, we only need to count the number of ”corners” and then fit the machine learning model to get the correct classification (Figure 3.6). The machine learning model can be simply understood as a high-dimensional function, the features are the variables input to the function, and the weight [w1, w2, ..., wn] of the function is to obtain the optimal global solution by using the optimization method through the historical data. After the model is trained, we input the extracted features [x1, x2, ..., xn] into the model y = w1*1 + w2 *2 + ... + wn*n to get the predicted classification Result y. Here we take the simplest linear classifier as an example.

However, the problems faced by computer vision are far more complicated than the above problems. For a long time, academia and industry have worked hard to design better ”features” and ”classifiers.” Such as the HOG feature describing the texture feature and the SVM algorithm that can solve linear and nonlinear problems through the kernel. However, due to the complexity of the real world, it is difficult for artificial experience to correctly and accurately transform perceptual cognition into digital features to describe. Since the machine learning

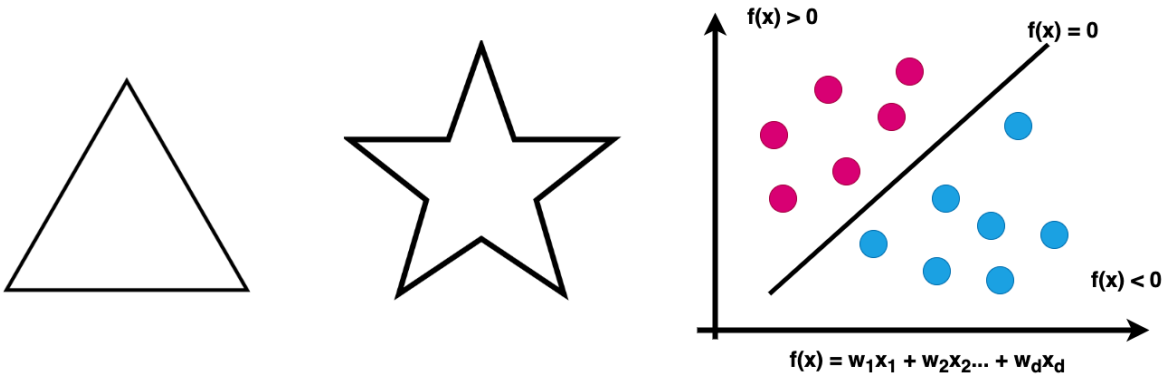

Figure 3.6: Object Feature Extraction.

model can fit a classifier to distinguish categories through data, is it possible to use data to extract features that indicate objects' essence? According to this idea, experts and scholars have proposed deep learning algorithms.

3.3.4.5 Deep Learning Algorithm Ideas

At the beginning of deep learning, it mainly played the role of feature extraction. During 2014, we often saw the two-stage training mode of classifiers, such as deep learning + SVM. This approach is very intuitive, but there are cases where the learned features need to match the desired task. With the algorithm's evolution, people gradually found that using the deep learning end-to-end model can achieve better results. The essence of image features is abstract painting information that can highly summarize the image content. The features extracted by artificial experience are to extract texture, color, shape, and other information from the original pixels from the human perspective. These features are derived from the actual pixels of the image. Although this is an efficient way to abstract information, it will lose a lot of data simultaneously. Since the artificially extracted feature source is calculated based on the original image, there is no information loss when using the model to learn directly from the original data. However, the valuable information contained in the image is very sparse, and the model will fail to converge if it pays too much attention to useless information. The proposal of a convolutional neural network solves this problem very well. Convolution is a concept in analytical mathematics. The easy-to-understand explanation of convolution is to extract the

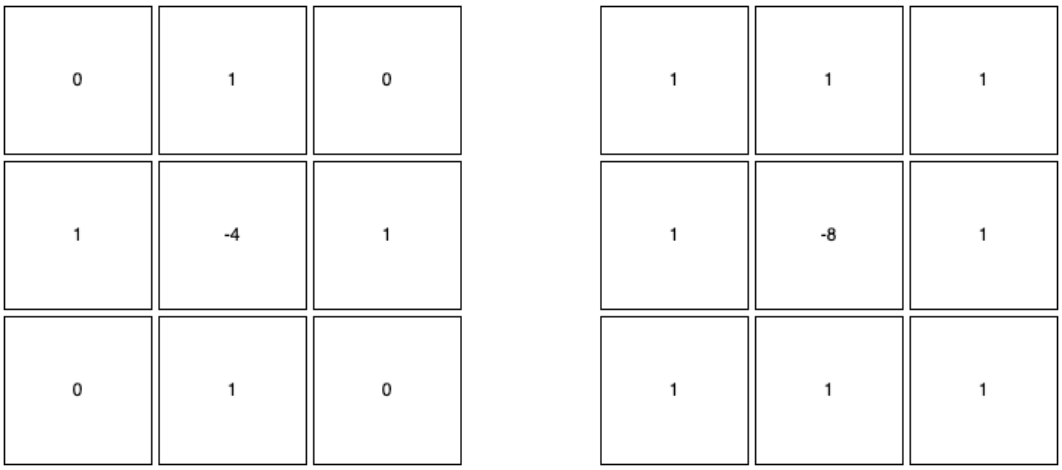

Figure 3.7: Laplacian filters

information we want from the original data by weighting the original data with the convolution kernel of the preset weight. In image processing, different convolution kernels are often used to process images, such as the Laplacian operator for extracting edge information.

The convolution kernel parameters in the above example (Figure 3.7), namely [0, 1, 0, 1, -4, 1, 0, 1, 0] are set through manual experience, it has strong interpretability, and it proves that The rationality of this setting method. The convolutional neural network allows the algorithm model to automatically learn the parameters of the convolution kernel in the labeled data, so that the algorithm can extract useful features based on a specific task without human intervention. The original convolutional neural network LeNet is a network model composed of convolutional layers, pooling layers, and fully connected layers. The main function of the convolutional layer is to extract image features. The role of the pooling layer is to reduce the amount of calculation while preserving key information as much as possible. The fully connected layer plays the role of a classifier, and cooperates with the activation function of nonlinear transformation. The minimum of a CNN The system is formed. Although there are many evolutions and variants of CNN in the follow-up research, the core ideas and components are composed of the above-mentioned parts. After the success of LeNet, theoretically, the more convolution kernels, the more features can be extracted. Therefore, a large number of researchers began to stack (deepen and widen) convolution kernels in pursuit of better algorithm effects, but the experiment did not achieve the expected results. . In order to improve the effect

of the algorithm, the CNN model has derived several schools. The convolution kernel parameters in the above example, namely [0, 1, 0, 1, -4, 1, 0, 1, 0], are set through manual experience. It has strong interpretability and proves the rationality of this setting method. The convolutional neural network allows the algorithm model to automatically learn the parameters of the convolution kernel in the labeled data so that the algorithm can extract useful features based on a specific task without human intervention. The original convolutional neural network LeNet is a network model composed of convolutional layers, pooling layers, and fully connected layers. The main function of the convolutional layer is to extract image features. The role of the pooling layer is to reduce the amount of calculation while preserving critical information as much as possible. The fully connected layer plays the role of a classifier and cooperates with the activation function of nonlinear transformation. The minimum of a CNN The system is formed. Although there are many evolutions and variants of CNN in the follow-up research, the core ideas and components are composed of the above-mentioned parts. After the success of LeNet, theoretically, the more convolution kernels, the more features can be extracted. Therefore, a large number of researchers began to stack (deepen and widen) convolution kernels in pursuit of better algorithm effects, but the experiment did not achieve the expected results. To improve the effectiveness of the algorithm, the CNN model has derived several major genres ResNet genre In theory, the increase in the convolutional layers will improve the effect, but the experiment found that the results are not as expected. When the number of convolutional layers increases to a certain level, the impact on the training set decreases instead (not caused by overfitting, the phenomenon of overfitting is that a high accuracy rate can be achieved on the training set and the accuracy rate on the test set decreases). This problem is because when the convolutional neural network is trained, the chain rule is used for backpropagation. When there are too many convolutional layers, the backpropagation gradient is prone to disappear or explode. ResNet is a network structure that can effectively solve the gradient disappearance/explosion proposed by Microsoft. Compared with simple convolution stacking, ResNet adds ”shortcut connections” to each block. This method (Figure 3.8) can directly transfer the gradient of the block input to the block output so that regardless of whether the convolution operation in the block will cause the gradient to disappear, the original gradient will be saved and passed down. The shortcut

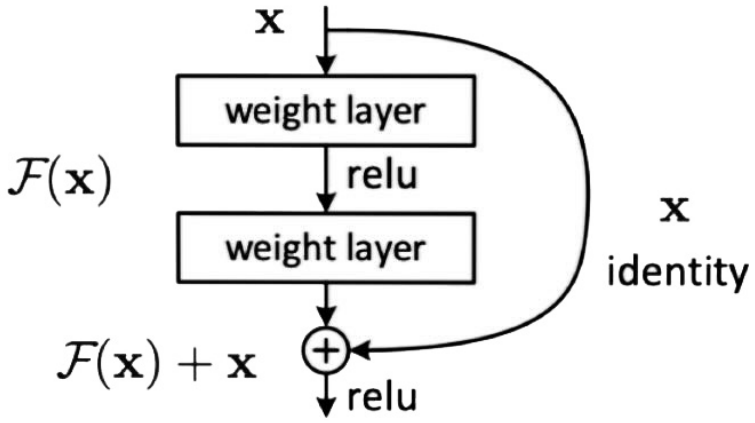

Figure 3.8: ResNet Structure

connection usually uses the identity method (The weight is 1). If the weight is greater than one or less than one, the problem of gradient explosion or disappearance will theoretically occur. Because the gradient will not disappear or explode, the network layers of ResNet can be made very deep. The commonly used ones are resnet18, resnet50, and resnet101, which can be selected according to the limitation of computing power.

Many scholars have interpreted the ResNet structure. There is a fascinating explanation that the ”shortcut connections” structure of ResNet can be expanded like a parallel circuit. After expansion, ResNet does not increase the depth of the network but increases the width of the network.

Inception genre Inception is another important genre proposed by Google. There are many versions of this genre, basically mutating and evolving yearly. The essential idea is to use convolution kernels of different kernel sizes to widen the network. Using convolution kernels of different sizes can avoid the problem of making the network too deep to increase the model's receptive field. Figure 3.9 shows a very classic block structure of inception v1.

There are many scholars who have interpreted the success of this structure. There is a more interesting point of view that the 1x1 convolution in inception plays the role of ”shortcut connection” in ResNet. However, ResNet is a more complex model with a higher number

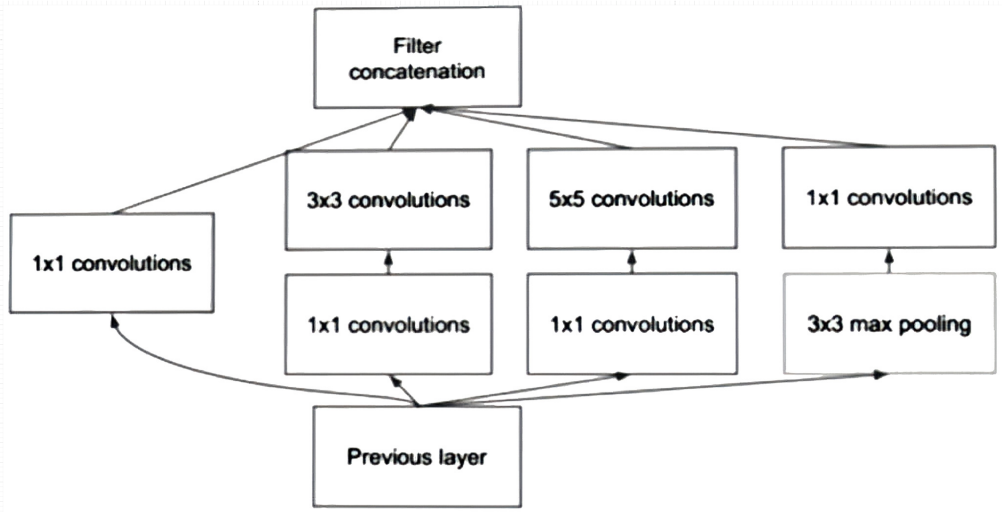

Figure 3.9: A classic block structure of inception v1

of parameters, making it more accurate but also more resource-intensive. This makes it less suitable for deployment on mobile devices with limited resources.

Mobile devices have limited computational resources and storage capacity, making it challenging to deploy deep learning models on them.

MobileNet SSD, a real-time object detection model that combines the MobileNet architecture with a Single Shot Detector (SSD), addresses these challenges by using depthwise separable convolutions, which reduce the number of parameters and computations required while maintaining high accuracy in object detection. It is trained on a large dataset of annotated images, which allows it to recognize and classify objects in real-time video streams. The model uses a sliding window approach to generate a set of candidate regions, which are then classified using a set of convolutional neural networks. The output of the model includes the location and class of the detected objects. The MobileNet architecture is the backbone of the MobileNet-SSD model, providing the feature extraction capabilities needed for object detection. The MobileNet-SSD model uses a modified version of the MobileNet architecture, including additional convolutional layers and feature maps, to enable object detection. The SSD algorithm is used to generate a set of candidate regions for object detection. Its algorithm uses a set of predefined aspect ratios and scales to generate a set of anchor boxes at each location in

the feature map. The anchor boxes are used to predict the location and class of objects in the image.

Furthermore, MobileNet SSD has been optimized for mobile devices, with efficient implementations available for mobile platforms such as Android and iOS. This ensures that it can be deployed and run efficiently on mobile devices without compromising performance. In addition to being efficient and fast, MobileNet SSD is also accurate in object detection. In the COCO dataset, MobileNet SSD achieved an mAP of 21.7 on a mobile device, comparable to the mAP of 21.9 performed by ResNet on a high-performance computing device.

Meanwhile, MobileNet SSD was six times faster than Faster R-CNN on a mobile device. And it outperformed YOLOv2 in terms of speed and accuracy on a mobile device. Therefore, MobileNet SSD is preferred over ResNet, Faster R-CNN, and YOLOv2 for object detection on mobile devices due to its smaller size, faster inference speed, and efficient implementation for mobile platforms.

3.3.5 Our Approach

For the NAAD project, we chose the iPhone 12 Pro max and Ipad Pro as the mobile devices equipped with a LiDAR scanner, which has a highly accurate distance measurement capability compared to other phones with depth maps. People with impaired vision will gain better spatial awareness. The application system first uses LiDAR technology to scan the environment around the user in real-time and builds a virtual 3D world through AR (ARKit) technology. Next, the system converts video of the user's surroundings captured by the camera on the mobile device into a sequence of images. According to the user's voice command, the system uses the TensorFlow Lite framework with the Mobile-Net Single-shot Detection deep learning model to analyze the image sequence in real-time and view the area of the target object on the screen of the mobile device. After that, to locate the target Object, the system samples the center point of the selected target object area and performs 2D to 3D coordinate transformation. The transformed 3D coordinates are the final position of the target object positioned by the system. For the object detection module, it includes target detection and distance estimation. The function of target detection is accomplished by the Machine Learning module. First, the

real-time video is transmitted to the target detection neural network in the form of frames. The final output of the model is the type of object and the bounding box. The function of distance estimation is accomplished by the AR Module. According to the 2D coordinates of the bounding box outputted by the neural network, we get the central point of the bounding box as the input data of the AR Module and then use the central point of the bounding box as the starting point to launch a ray into 3D space. When the Ray Collides with the object at the point of impact, we can calculate the distance from user A (x1, y1, z1) to object B (x2, y2, z2) based on the EUCLIDEAN distance. This distance is then returned to the main logical core module, which continues to execute the user instructions in the corresponding mode, and finally presents the data to the user in visual, tactile, and acoustic form. This can greatly improve the user's ability to identify obstacles in the surrounding environment and reduce the risk of users walking indoors or outdoors. Using the right vector of the camera and the vector of the camera to the target object in the AR system, we can get the angle of the target object relative to the user. The angle range is divided into three categories: left, right, forward. For Distance Setting, we have three levels: normal distance (2m $\leq$ d $\leq$ 5m), safe distance (1m $<$d $\leq$ 2m) and warning distance (d $\leq$ 1m).

3.3.6 Interface design

In the design and development of the NAAD system, we chose to use Unity to provide an intuitive visual programming interface for various development tasks. Figure 3.10 shows the user interface in the NAAD system created with Unity. To allow users to use this system more conveniently, we use large-area and strong-contrast click buttons with bold red letters on a black background. It can make it easier for VIB people to operate the system. Users can enter the NAAD system's query mode or security mode by voice command or by tapping the query button at the bottom right of the screen or the security button at the bottom left of the screen, respectively. When the user enters the query mode, as shown in Figure 3.10, the target object is the TV that the user searches through the voice or gesture commands. The system will automatically detect and mark the target object. Then the system will broadcast the object's category, distance, and orientation in real-time and display the detection distance

Figure 3.10: Query Mode in NAAD.

(4.22m) below the main interface. If the user clicks the microphone button in the lower right corner of the interface, the system will receive the user's new voice command and perform corresponding operations. If the user clicks the speaker button in the lower left corner of the interface, the system will broadcast the relevant information about the target object to the user again.

3.4 Safe mode in NAAD system

3.4.1 Research Problem

Blind and visually impaired people cannot accurately judge the changes in their surroundings due to their eyesight limitations, which leads to the risk of accidents when they walk outdoors or indoors.

3.4.2 Research Questions

- How to notify the user of the nearest obstacle in real-time?

- How to improve the distance detection accuracy between the obstacle and the user?

3.4.3 Research Hypothesis

- The NAAD system can quickly and effectively assist visually impaired and blind (VIB) people in avoiding obstacles and reducing the safety risks in daily life.
- The Safety Mode of the NAAD system can notify users of the category, location, and distance of the nearest obstacle in real-time.
- The NAAD system can provide an accurate distance detection function within a specific range.

3.4.4 Obstacle detection

3.4.4.1 Introduction

In daily life, visually impaired and blind people often face the safety problem of how to avoid obstacles in time due to their lack of perception of visual information. It limits their ability to process their surroundings and interact with society, hindering their daily activities and reducing their quality of life. Over the years, obstacle avoidance assistance techniques for the VIB population have evolved, and researchers worldwide have developed various methods to help these individuals detect and avoid obstacles. In this chapter, a comprehensive overview of different obstacle detection and avoidance methods is given, and their effectiveness is compared. The focus is on methods that combine AR and machine learning. Among these methods, the combination of augmented reality (AR) and machine learning has shown great potential to solve this problem. However, using current assistive technologies poses challenges, as aspects of flexibility and portability remain an issue due to hardware and usability constraints. In this paper, we develop an application that satisfies both cognition and spatial awareness - NAAD - on a LiDAR-enabled mobile device. First, it utilizes deep learning and LiDAR-based AR technology to detect obstacles and calculate the distance between them and the user. Then the system will issue a voice prompt message for the user according to the collected data information. The voice prompt information includes the nearest obstacle category as well as the

direction and distance. The system has also developed an inclusive user interface for VIB people, which can easily complete functions such as obstacle detection, distance estimation, and obstacle reminder by using both gestures and a voice command feature.

3.4.4.2 Problem overview

There are many traditional obstacle avoidance aids and methods, such as glasses, magnifying glasses, canes, guide dogs, etc., which are used by visually impaired and blind people because they can help them complete basic daily tasks such as obstacle avoidance. Among them, glasses and magnifying glasses can only help people with mild visual impairment. They can't do anything for people who are severely visually impaired or blind. The white cane is the most traditional obstacle detection and avoidance method for the visually impaired. The cane is placed in front of the user and taps the ground to detect any obstacles. Based on feedback from the cane, the user adjusts the path to avoid obstacles. Canes are very effective for detecting unusual obstacles on the ground. However, when VIB people use a cane, they need to constantly repeat the detection, and it is not effective for detecting obstacles above the waist. Guide dogs are trained to help the visually impaired find and avoid obstacles. The dogs use their senses of smell, hearing, and vision to detect and avoid obstacles and are trained to respond to specific commands from users to get around obstacles. However, at least 2.2 billion people worldwide have near or distance vision impairment and the number of qualified guide dogs worldwide is far less than the demand. And the training cost of a guide dog is as high as 25,000-3,000 US dollars, the training period is as long as 18 months, and the working life is only 8-10 years. And many public places still prohibit the entry of guide dogs, which also brings great trouble to the lives of blind people. From the perspective of guide dogs, they also sacrifice their playful nature because they must resist external interference and temptation when working. In addition, although guide dogs can detect obstacles, their communication with human family members is often unclear. VIB personnel cannot know the size, height, and danger of specific obstacles, so they cannot effectively help VIB personnel avoid obstacles. Today, with the development of technology, these devices can be further expanded by exploring and applying intelligent sensing technology, augmented reality (AR) technology, and machine learning technology.

3.4.4.3 Related study

Obstacle avoidance and navigation are the most difficult tasks in assisting VIB people in their daily activities. Obstacle avoidance is the basic requirement for autonomous navigation, and autonomous navigation technology is an important symbol of modern intelligent navigation. With the rapid development of science and technology, the disadvantages of traditional obstacle avoidance methods are becoming more and more obvious. To address these challenges, experts have developed obstacle-avoidance navigation systems such as e-navigation and various smart glasses that utilize a combination of AR and machine learning. Electronic navigation systems are devices that use sensors to detect obstacles and provide audio feedback to the user. The system can use sonar, ultrasound, or infrared to detect obstacles, and audio feedback informs the user of the location and distance of the obstacle. Smart glasses are wearable devices that use cameras, sensors, and artificial intelligence to detect obstacles and provide audio feedback to the user. The smart glasses can detect obstacles in real time, and the user receives audio feedback about the location and distance of the obstacle. However, electronic navigation systems and smart glasses both use sensors for obstacle avoidance and audio feedback to help individuals detect and avoid obstacles. Intelligent sensor obstacle avoidance refers to the technology or system that uses sensors to detect obstacles, and then uses artificial intelligence algorithms to process the information to make obstacle avoidance decisions. The goal is to make the robot, self-driving car or blind person perceive static or dynamic objects that hinder their movement through sensors, update the real-time path according to a certain algorithm, avoid obstacles, and finally reach the destination. Sensors can include cameras, ultrasonic, infrared, LiDAR sensors, and more. AI algorithms can use techniques such as computer vision, machine learning, or path planning to make obstacle avoidance decisions.

In the autonomous movement of people, the sensor plays a very important role. It can perceive the surrounding environment information in real time, including the size, shape, position, posture, etc. of obstacles. There are many types of sensors used for obstacle avoidance, all of which have different characteristics. And the principle, which will involve a variety of sensors such as vision, ultrasonic, infrared, and laser radar.

Figure 3.11: Visual sensors

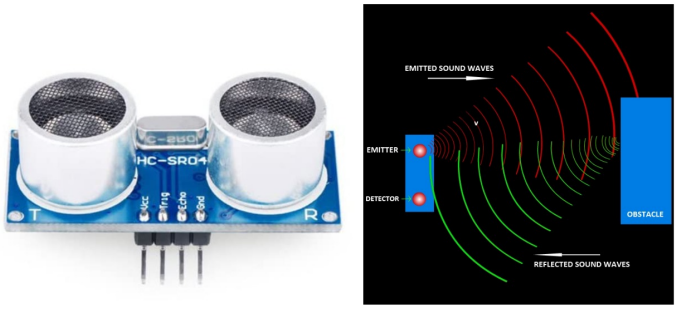

Figure 3.12: Ultrasonic Sensor

Visual sensor: mainly use monocular (Figure 3.11 left), binocular camera (Figure 3.11 right), depth camera, video signal digitization equipment or DSP-based fast signal processor and other external devices to acquire images, then perform optical processing on the surrounding environment, and process the collected image information Compression is fed back to the learning subsystem composed of neural network and statistical methods, and then the subsystem connects the collected image information with the actual position of the person to complete the positioning. It has the advantages of simple structure, multiple installation methods, no sensor detection distance limitation, and low cost, but it is greatly affected by ambient light and cannot work in dark places (non-textured areas).

Ultrasonic sensor: Ultrasonic sensors (Figure 3.12) can detect transparent materials such as glass and mirrors. It mainly emits ultrasonic waves through the transmitting probe. The ultrasonic waves encounter obstacles in the medium and return to the receiving device. By receiving the ultrasonic reflection signals emitted by itself, the propagation distance is calculated according to the time difference between ultrasonic emission and echo reception and the

Figure 3.13: Infrared Sensor

propagation speed. The distance between obstacles and people. However, since the speed of ultrasonic waves in the air is related to temperature and humidity, changes in temperature and humidity and other factors need to be taken into account in more accurate measurements. Ultrasonic sensors are also affected by the surrounding environment (mirror emission or limited beam angle) and the external environment such as obstacle shadows and rough surfaces, and the applicable range is small and the ranging distance is short.

Infrared Sensors: Most infrared sensor (Figure 3.13) ranging is based on the principle of triangulation. The infrared emitter emits an infrared beam at a certain angle, and when it encounters an object, the beam will be reflected back. After the reflected infrared light is detected by the CCD detector, an offset value L will be obtained. Using the triangular relationship, after knowing the emission angle, the offset distance L, the central moment X, and the focal length f of the filter, the sensor The distance D to the object can be calculated through geometric relations. The advantages of infrared sensors are that they are not affected by visible light, can be measured day and night, have high angular sensitivity, simple structure, wide measurement range, long distance measurement, and short response time. However, it is greatly affected by the environment, and the color, direction, and surrounding light of the object can all cause measurement errors, and the measurement is not accurate enough. In addition, the infrared sensor cannot detect the distance to black objects or transparent objects, and cannot work normally in the case of occluded objects.

Figure 3.14: Lidar Sensors

Although there are many sensors on the market, lidar (Figure 3.14) is still the most mature core sensor in assisting human navigation and obstacle avoidance. Lidar (Light Detection for Laser Imaging and Ranging) technology is a typical remote sensing technology. It can accurately perceive the three-dimensional information of the surrounding environment and detect the precise position of the object, and the detection accuracy is within the centimeter level. This enables lidar to accurately identify the specific outline and distance of obstacles without missing or misjudging obstacles ahead. How to apply lidar to smartphones? We can't directly put lidar equipment on the unmanned vehicle on the mobile phone, because the cost is too expensive. Apple uses Flash lidar. Although the detection distance of Flash lidar has not yet reached the requirements of self-driving cars, the cost is very low. In the lidar introduction on Apple's official website, it is indicated that "lidar scanning can measure objects 5 meters away, and can work at nanoscale speeds both indoors and outdoors." Under the condition of limited distance, flash lidar has the strength to "contempt" other lidars in terms of stability, cost, and measurement accuracy. Apple's strategy is Flash+DTOF.

According to different ranging principles, lidar can be divided into time-of-flight (dTOF) lidar and phase offset (iTOF) lidar. Direct calculation of distance is the principle mentioned at the beginning of this article, and it is also the most commonly used ranging principle for vehicle-mounted lidar. As shown in Figure 3.15, dToF directly measures time-of-flight. The principle is to use nanosecond or even picosecond-level short-pulse laser directly at the transmitting end, and react quickly after emission, and quickly receive the reflected laser light. On dToF, it calculates the distance by recording the time interval between the transmitted pulse and the received pulse. dToF will transmit and receive light signals N times within a single

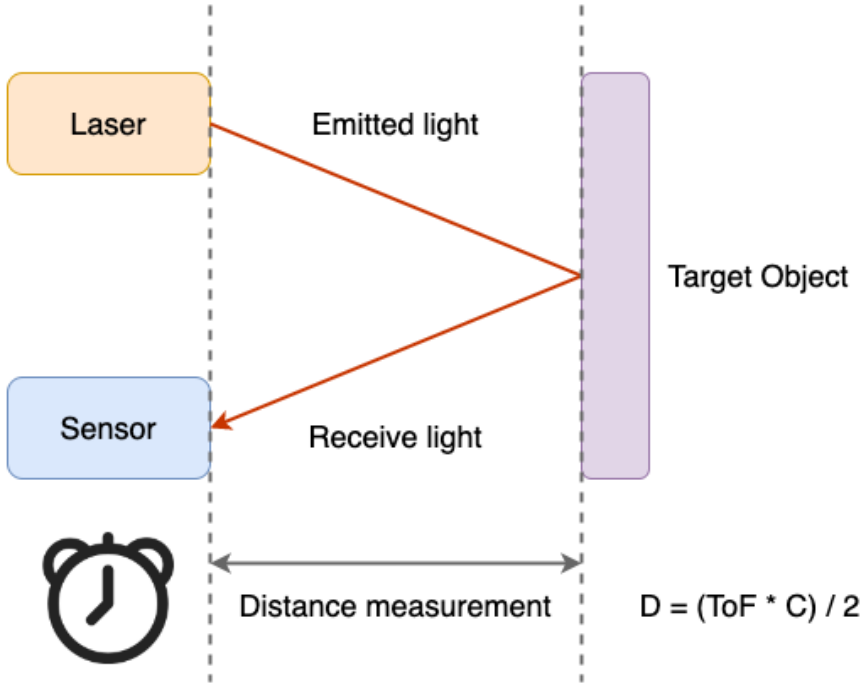

Figure 3.15: Light and Time-of-Flight (ToF)

frame measurement time, and then analyze and count these times, and obtain the final distance. But at this stage, due to the need to make a good synchronization judgment on the flight time and receiving time of the optical pulse, the requirements for the algorithm and hardware are relatively high. Previously, only most high-end cars used dToF until Apple applied this technology to iPads and iPhones. Phase-offset (iTOF) LiDAR uses modulated laser light emitting at a specific frequency to calculate distance indirectly by calculating the phase difference. iTOF has low precision, weak anti-interference, and high power consumption, but the process is relatively simple.

3.4.5 Our Approach

In terms of power consumption, the pulse wave emission adopted by dToF has a lower duty cycle than the continuous wave of iToF, and can emit more targeted light sources within the same time. Compared with the power consumption of dToF It will be smaller and more suitable for use on devices with less power. From the perspective of application scenarios, dToF has low power consumption, small size and is suitable for use in smaller devices, and because of

its better immunity to interference, it is also better for outdoor use. And due to the principle of dToF, the accuracy will not be greatly reduced when the measurement distance increases, and the energy consumption will not be greatly increased. For example, the use of AR in the direction of Apple's choice is a good development direction. For example, Apple products such as the iPhone 12 Pro and Pro Max, iPhone 13 Pro and Pro Max, and iPad Pro now have built-in lidar scanners that can measure the distance of surrounding objects up to 5 meters away, and can be used both indoors and outdoors. Outdoors and operating at the photon level at nanosecond speeds. People can obtain the dot matrix data of the surrounding environment through real-time scanning of lidar. Combined with the corresponding SLAM algorithm, people can realize intelligent navigation in an unknown environment (no need to enter the map in advance), judge the best path to the destination, avoid obstacles, and reach the destination smoothly. When developing the NAAD system, we chose iPhone 12 Pro Max and iPad Pro as development devices, using lidar combined with deep learning and AR technology to quickly and accurately identify obstacles and measure distances, so as to achieve the function of prompting VIB people to avoid obstacles in real time.

Figure 3.16 illustrates the NAAD system workflow. Users interact with the virtual assistant through a mobile device. The virtual assistant utilizes deep learning, AR, and Lidar sensor to provide feedback information utilizing sound, vision, and vibrations to indicate the presence of obstacles and assist a user in avoiding objects. The user can use this information to avoid obstacles and to assist them in finding items that they are seeking. The NAAD system aims to provide a safer environment for users by detecting obstacles and estimating the distance between the user and them. We developed NAAD system as an Apple app, the iPhone 12 Pro Max was chosen as the mobile device equipped with a LiDAR scanner. LiDAR technology scans the user's surroundings in real time, constructs a virtual 3D world through augmented reality (ARKit) technology, and then converts the captured video into a series of images. The system uses the TensorFlow Lite framework, combined with the Mobile-Net Single-shot Detection deep learning model, analyzes the image sequence in real time, and locates the target object according to the user's voice command. The system first samples the center point of the selected target object area, locates the target object, and performs coordinate transformation from 2D

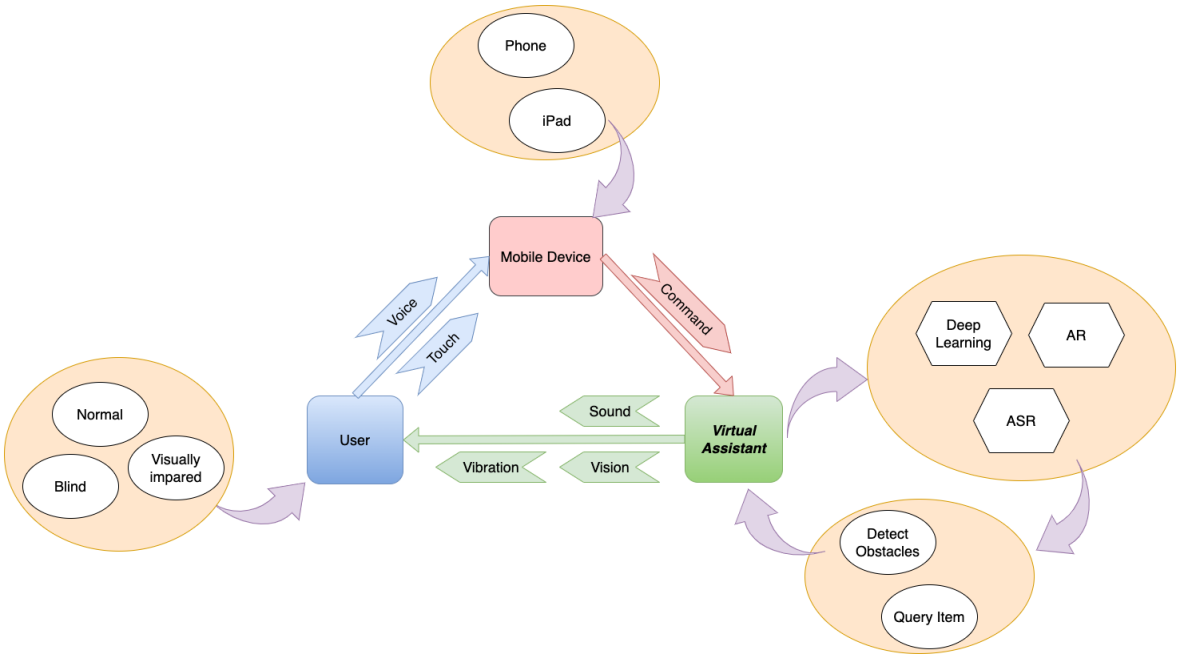

Figure 3.16: NAAD System Overview.

to 3D. The transformed 3D coordinates represent the final location of the target object. The safety modules of the NAAD system include obstacle detection and distance estimation. The machine learning module performs the obstacle detection function, where the real-time video is transmitted in the form of frames to the object detection neural network, and the final output of the model is the type and bounding box of the object. The AR module does the distance estimation function. Based on the 2D coordinates of the bounding box, the center point of the bounding box is used as the starting point to emit a ray to the 3D space. When the ray hits the object, the distance from the user to the object can be calculated. When the distance between the user and the obstacle is less than 1 meter (recommended safety distance, user can customize it). The angle of the target object relative to the user is also determined by the camera right vector and the camera to target object vector in the AR system. Angle range is divided into left, right, front three categories. In conclusion, the NAAD system provides users with a convenient and intuitive visual programming interface, which can significantly improve the user's ability to identify surrounding obstacles, reduce the risk of walking indoors or outdoors, and provide visual, tactile and auditory sensations to enhance users' awareness of the surrounding environment. perception of the environment.

Figure 3.17: Safe Mode in NAAD.

3.4.6 Interface design

Figure 3.17 shows the user interface in safe mode in the NAAD system created with Unity. To allow users to use the system more conveniently, we continue to use the large-area, contrasting click buttons with red letters on a black background. It can make the operating system easier for VIB people. When the user enters the safe mode of the NAAD system through voice or gesture commands, the system will alert the user to the nearest obstacle. Figure 3.17 shows that the system first recognizes the cup and the laptop. The system will then measure the distance between these two objects and the user and use a red crosshair to mark the closest laptop to the user as an obstacle. Finally, the system will voice prompt the user with the obstacle category, the distance, and the orientation between the obstacle and the user. Then, the system displays the detection distance (0.68m) at the bottom of the main interface. If the user clicks the microphone button in the lower right corner of the interface, the system will receive the user's new voice command and perform the corresponding operation. If the user clicks the

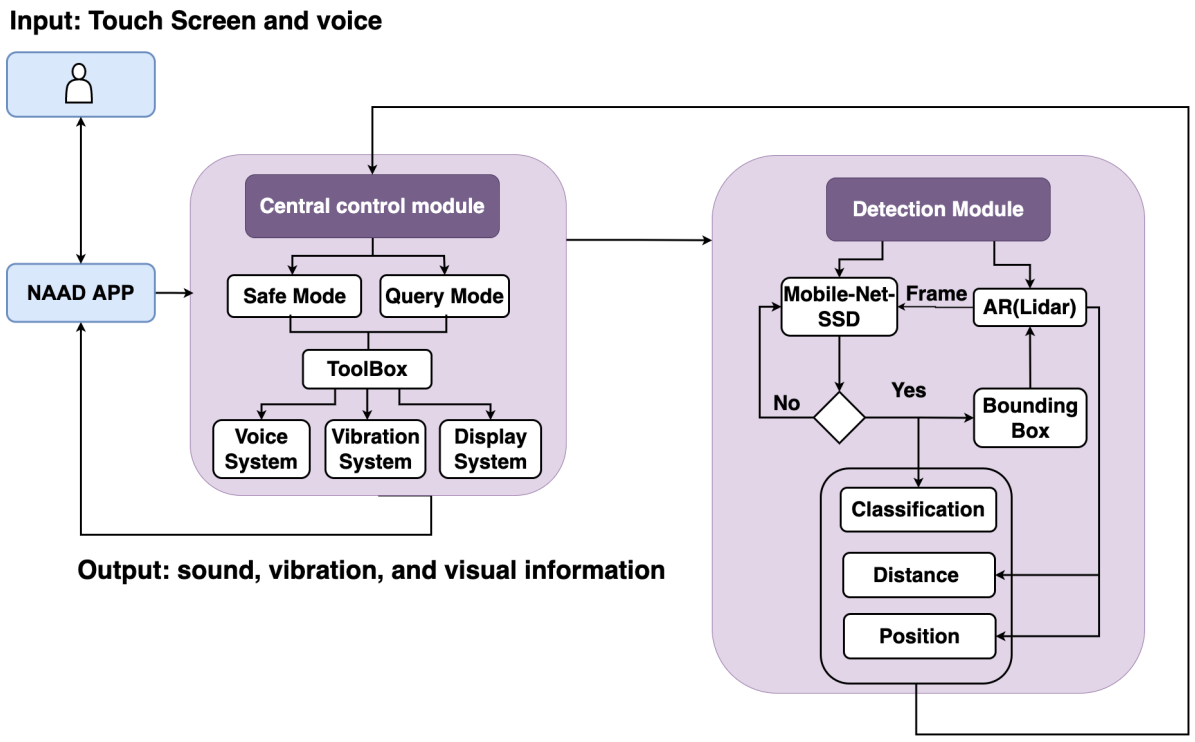

Figure 3.18: NAAD System Structure

speaker button in the lower left corner of the interface, the system will broadcast the relevant information about the obstacle to the user again.

3.5 NAAD system design

The NAAD system is mainly divided into the main logic core module and the object detection module that provides a virtual voice assistant to interact with users. Figure 3.18 describes the main logic core module and the object detection module.

The main logic core module is divided into security mode and search mode. In safe mode, the object detection module detects all obstacles smaller than 1 meter and notifies the user. In search mode, the user can find the name of the item they are looking for in a list of identifiable items provided by the system, and then tell the virtual voice assistant by speaking. The virtual voice assistant instructs the user on how to scan the environment. When it finds the object, it informs the user that the object has been found and tells the user the distance from the object

to the user and the direction of the object. In addition, the system provides a feedback manager, Toolbox, which consists of three subsystems: sound system, vibration system, and visual display system. The sound system includes two-way conversion from sound to text and from text to sound. For example, the system converts the user's voice into the text to the virtual voice assistant, who converts the feedback from text to sound to the user. The vibrating system provides the function of vibrating at different frequencies. The visual display system displays the name and bounding box of the target obstacle and the searched object to the user in 2D and adds a 3D instruction model at the center of the bounding box. In this way, the target object can be more prominently displayed on the screen of the mobile phone so that the user can identify it more clearly.

3.6 Results and Discussions

We used the pre-trained model of Mobile-Net-SSD, which trained on the COCO dataset, which contains 300k images, 1.5 million object instances, and 90 different categories of items. The input image size is 300*300*3. We integrated the model of Mobile-Net-SSD and AR system into unity. Also, we captured more than 50 images when testing the system and retrieved data to calculate the experiment results.

This experiment is divided into four groups. In the first group, the system estimates the accuracy of the user's distance from the object according to the different objects (Refrigerator, TV, and Chair).

$$(V_O - V_A)/V_A * 100 \tag{3.1}$$

For example, when a user is less than one meter from an object, the system estimates the distance values Vo between the user to the obstacle. We set the real distance is VA and use the equation (3.1) to calculate the percent accuracy of distance. We repeated the process at different distance levels. The result is shown in Table 3.1. The distance accuracy of the system is best between three and four meters. Since our experiments are not conducted in a simulated lab but in a real user's room, our average distance accuracy in complex environments is over

96%, indicating that our system can maintain high range accuracy within a five-meter indoor range.

Distance Accuracy	$0.5m<D \leq 1m$	$1m<D \leq 2m$	$2m<D \leq 3m$	$3m<D \leq 4m$	$4m<D \leq 5m$
Refrigerator	96%	94%	97%	98%	96%
TV	93%	97%	98%	99%	96%
Chair	92%	95%	96%	99%	97%

Table 3.1: Distance Accuracy

In the second set of experiments, the system's virtual assistant will give a speech prompt for the recognized obstacle. We tested the prompts for categories of items and the distance between the item and the user. If the prompt is correct, we mark it as YES, otherwise, we mark it as NO. For example, when the system detects a TV 2.1 meters in front of the user, the virtual assistant will tell the user that "the TV is more than two meters in front of you.". From Table 3.2 we could see that for these three common household appliances, the system's voice broadcast accuracy rate is 100%.

Sound Accuracy	$0.5m<D \leq 1m$	$1m<D \leq 2m$	$2m<D \leq 3m$	$3m<D \leq 4m$	$4m<D \leq 5m$
Refrigerator	Yes	Yes	Yes	Yes	Yes
TV	Yes	Yes	Yes	Yes	Yes
Chair	Yes	Yes	Yes	Yes	Yes

Table 3.2: Sound Accuracy

In the third experiment, we evaluated the reaction time of the system to identify the object. Robert Billers argued that, ideally, users should get feedback on their actions within 100 milliseconds, because the fastest subliminal movements are those in which the blink of an eye

lasts between 100 and 150 milliseconds, and 100 milliseconds feels like an instant. From Table 3.3 we derived an average reaction time of 19 milliseconds. This result can greatly improve user experience.

Response Time (ms)	$0.5m<D \leq 1m$	$1m<D \leq 2m$	$2m<D \leq 3m$	$3m<D \leq 4m$	$4m<D \leq 5m$
Refrigerator	20	20	20	19	19
TV	19	18	19	19	18
Chair	20	18	18	18	19

Table 3.3: Response Time (millisecond)

In the fourth experiment, we tested the accuracy of the system for object recognition. Depending on the range, we get an average object recognition accuracy of 73% within a five-meter range. There are two possible reasons for this. First, the lighting problem. Insufficient Indoor illumination will lead to a decrease in the object recognition rate. Although LiDAR is not affected by illumination, image recognition based on deep learning is affected by illumination. [32] proposed that the problem can be solved by using depth maps. Second, it has to do with the image of the training data. For example, if the images of the training data are taken within two meters, the recognition accuracy will be improved. From Table 3.4 we can see that the high precision is concentrated within two meters and the low precision is within two to five meters. We can increase the accuracy of object recognition by adding more extensive training data.

3.7 Conclusions

In this research, we proposed a visual obstacle recognition framework based on object detection and augmented reality (AR) for obstacle recognition and object retrieval. This framework has strong extensibility and generality. The modules in this framework are flexible, and the system implemented based on this framework can be applied to multiple platforms. Also, this system requires only one mobile phone without the need to carry additional detection equipment or

Object Detection Accuracy	$0.5m<D \leq 1m$	$1m<D \leq 2m$	$2m<D \leq 3m$	$3m<D \leq 4m$	$4m<D \leq 5m$
Refrigerator	73%	73%	77%	73%	76%
TV	75%	75%	68%	76%	58%
Chair	82%	82%	65%	75%	66%

Table 3.4: Object Detection Accuracy

any network requirements. It is far more portable than any other existing system. The system also provides a virtual assistant to interact with the user in real-time, which greatly simplifies the user's operation. For future work, we will improve the user interaction design of the virtual assistant and provide richer real-time feedback and more intelligent operation to help users quickly assess the surrounding environment to reduce the risk of unnecessary accidents.

Chapter 4

Optimized NAAD system with navigation functions

4.1 Introduction

Finding lost items presents a significant challenge for individuals with visual impairments daily. This challenge is compounded by their visual deficiencies, especially when they are in different places. Many navigation devices have been designed to assist the visually impaired in real life. But these devices require additional purchases, lack flexibility, and cause inconvenience to visually impaired users. To address these issues, we proposed the NAAD system that integrates LiDAR technology on mobile devices to enable obstacle avoidance, indoor navigation, and target object detection by integrating Deep Learning and AR technologies. Our innovative approach encompasses object recognition, route planning, and target object navigation to guide VIB users in locating specified lost items in the same or different spaces. This paper underscores the novel methodology employed in the NAAD application, which aims to enhance the quality of life for VIB people.

In accordance with the World Health Organization, a minimum of 2.2 billion individuals globally experience near or distant vision impairment. Visual impairment seriously affects their quality of life. Specifically, visually impaired people frequently encounter daily challenges, notably when searching for lost items indoors, resulting in heightened anxiety and depression due to an absence of visual information. As computer vision technology increasingly permeates society, visually impaired individuals may avail themselves of such technological advancements to gain more comprehensive visual information regarding their surrounding environments. Notably, numerous studies have established the capacity of computer vision technology to assist

VIB people with navigating [31] and identifying objects [32]. However, most assisting devices currently require additional purchases [33]. In addition, many assisting devices need to input a large amount of item information before searching for target items. These operations could be more convenient for the VIB people. Also, many assisting devices [34] for identifying objects cannot search for target objects that are far away or not in the same space as the user. To address these issues, we developed an innovative approach encompassing a design, programming, and interface, culminating in the NAAD system, which stands for Navigation Assistance through AR Technology and Deep Learning. This NAAD system can identify and alert users to avoid obstacles, enables the search for target objects, and navigation to assist VIB people in finding lost items in different spaces.

4.2 Research Problem

Finding lost items can be a significant challenge for visually impaired individuals because their vision problems make it difficult to locate and identify objects. This can result in frustration, decreased independence, and lost time and productivity. For example, a visually impaired individual might have difficulty finding their keys or phone, resulting in delayed departures or needing assistance from someone else. Additionally, the inability to quickly and easily find lost items can lead to anxiety, stress, and increased dependence on others. Therefore, improving the ability of visually impaired individuals to find lost items could be the development of new technologies that use alternative sensory inputs (such as touch, sound, or vibration) to help locate and identify lost items. Another potential research problem could be the study of how visually impaired individuals currently navigate and find lost items, in order to better understand the challenges they face and to identify areas where technology or other interventions could be most helpful.

4.3 Research Questions

- Is there an efficient way to to help visually impaired people find lost items in different spaces?

- How to enable a system to navigate the visually impaied people to the target object?

4.4 Research Hypothesis

- The NAAD system can easily and effectively guide users to find user-specified lost items in the same or different spaces
- The NAAD system can provide the shortest and safest path in different spaces.

4.5 Our Approach

In this work, we show a design of a portable, flexible, and easy-to-use indoor navigation application system to help visually impaired people quickly find lost items in different spaces. For this project, we chose the iPhone 12 Pro Max as a mobile device equipped with a LiDAR scanner, which has a highly accurate distance measurement capability compared to other smartphones with depth maps.

4.5.1 Detect and Locate Object

The NAAD system proposed in this paper utilizes LiDAR technology to scan the user's surroundings in real-time and creates a virtual 3D world by integrating ARKit technology. To identify the target object, the system captures video data of the user's environment via the mobile device's camera and generates an image sequence. Upon receiving the user's vocal command, the system employs the TensorFlow Lite framework with a Mobile-Net Single-shot Detection deep learning model for real-time image sequence analysis of the target object's area on the mobile device's screen. To pinpoint the target object's location, the system first samples the selected target area's central point and then converts the resulting 2D coordinates into 3D coordinates, which serve as the final position of the target object, as located by the system.

4.5.2 Guide The User to The Destination

In the initial phase of our project [35], we successfully identified, localized, and enabled users to find a target item within a 5-meter range. However, beyond this range or in different spaces, the system can not accurately identify and locate the target item. In order to overcome this

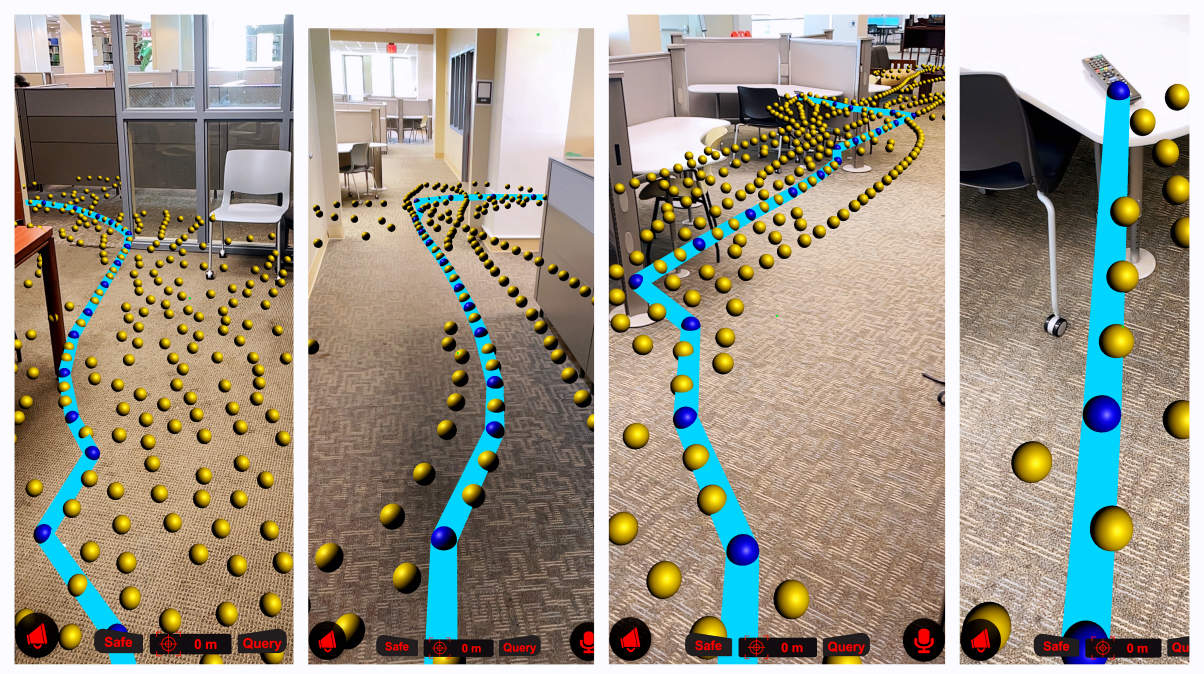

Figure 4.1: The NAAD system generates user's valid Position nodes (yellow spheres) and shortest safe navigation paths (blue trajectory) to guide the user to find the target object (remote control) in real-time.

issue, we introduced a memory storage unit in query mode, as depicted in Figure 4.2, which can record the real-time position of the user and the target item separately. The system generates new safe Position nodes according to the user's activities. These new Position nodes will form a dynamic location graph. The system automatically notes the item's category and Position information when the user interacts with a target item for the first time and adds them to the dynamic location graph. Based on this graph, we utilize Dijkstra's algorithm [36] to find the shortest safe path from the user to the target item. Therefore, when the user searches for the target item using voice commands, the system calculates the offset angle according to the user and navigation path directions.

Next, the system will automatically generate the shortest safe path and guide the user to the Position node of the target object. Finally, the system will prompt the user for all navigation information (including the name of the target object) through voice (front, left, right), somatosensory vibration, and eye catching visual signs to help the user find the target object. Due to mobile devices' limited memory and computing power, we cannot allow the number of

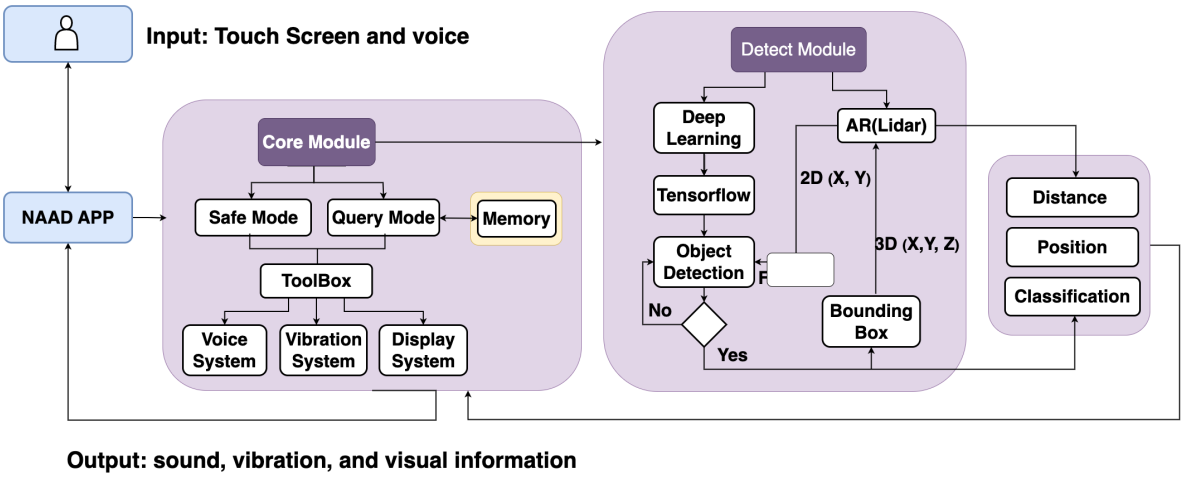

Figure 4.2: Architecture of the NAAD Application System.

new Position nodes to increase indefinitely. Therefore, the system we designed compares the distance between a new Position node and the generated Position node in the dynamic position graph to decide whether to add this new Position node. Users can also use voice commands to customize the distance between Position nodes based on the complexity of the space. For example, suppose the distance of the new Position node to the existing Position node is less than the distance of the user-defined Position node. In that case, the dynamic location graph will not add this new Position node. As shown in Figure 4.1, when the user is in the 2b2b area of 100 square meters, and the distance between the Position nodes is set as 0.2m, the system converts the user's activity trajectory into the safe Position nodes (yellow sphere). Through Dijkstra's algorithm, the system obtains the real-time optimal safety path (blue path) between the user Position node and the target object Position node.

4.6 System design

Figure 4.2 shows the architecture of the NAAD System. The system proposed comprises two major modules: an object detection module and a central control module. The object detection module features target recognition and distance detection capabilities. Specifically, the target recognition function receives frame pictures of the user's surrounding environment through the augmented reality (AR) system and feeds them to a neural network model (Mobile-Net-SSD).

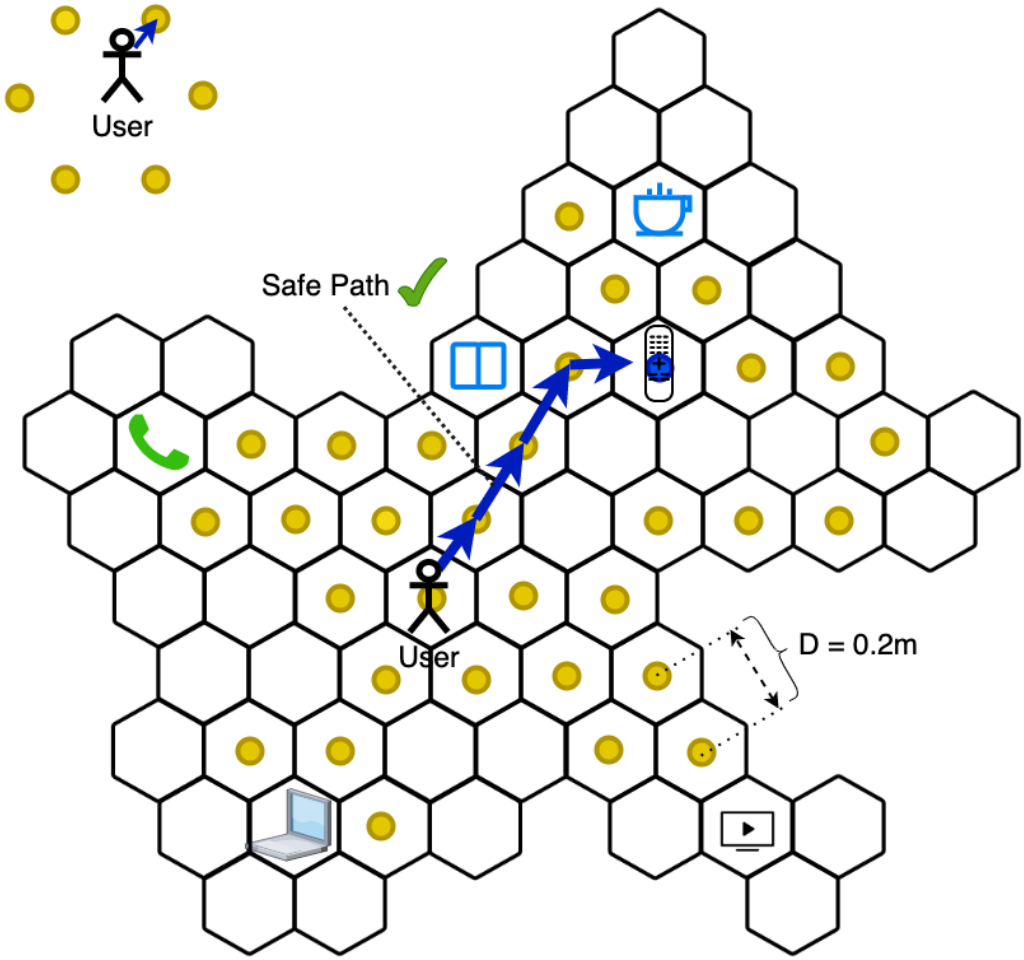

Figure 4.3: Planar Structure of The Memory Storage Unit.

The model produces output in the form of the item category and position coordinates in the bounding box. The AR system completes the distance detection function, which utilizes Raycast to calculate the distance between the user and the target object using the center point of the bounding box coordinates. The calculated distance is then communicated to the central control module. On the other hand, the central control module consists of two distinct modes: safe mode and search mode. In the safety mode, the central control module retrieves all information pertaining to the items detected by the object detection module based on the user-defined safety distance. It then issues real-time sound and vibration reminders, notifying the user about the category and position of the nearest obstacle. In the search mode, the user gives voice

commands to the central control module to locate the desired target item. If the position information of the target item is not available in the memory storage unit, the system utilizes the object detection module to scan the user's surrounding environment to locate the target item. If the user enters query mode and requests the target object, the system provides the distance and orientation of the target object to the user and records its position information in the memory storage unit for the subsequent generation of the navigation path. Conversely, if the position information of the target item is available in the memory storage unit, the system generates a safe path to guide the user to the target item automatically.

Figure 4.3 shows the planar structure of a memory storage unit. The memory storage unit only stores blue and yellow nodes. The blue node represents the target object which contains the location and category information of the target object. The yellow nodes represent the location points that the user walked through. We set the distance D between yellow nodes to a threshold of 0.2m. D can be adjusted according to the actual indoor area. It cannot be too small because mobile devices' memory and computing power are limited, and we cannot allow the number of new location nodes to increase infinitely.

The system will connect all yellow and blue nodes to form a network structure graph (see figure 4.4) stored in the memory unit. When the user queries the target object, the system will find the yellow node (see figure 4.3 top left corner) closest to the user in the network graph stored in the memory unit. And then, start from the yellow node, which is the closest to the user, and the blue node (target object) point as the endpoint to search for the shortest safe path. In the figure, the user finally found the remote control of the target object according to the blue shortest safe path generated by the system.

In addition, the central control module contains a Toolbox that consists of three subsystems: voice system, vibration system, and visual display system. The voice system includes a two-way conversion of sound to text. The vibration system occurs when the distance between the obstacle and the user is less than the warning distance in safe mode. The visual display system mixes 2D and 3D elements and highlights the name and bounding box of the target obstacle or searched object. It allows the user to identify the target object more clearly.

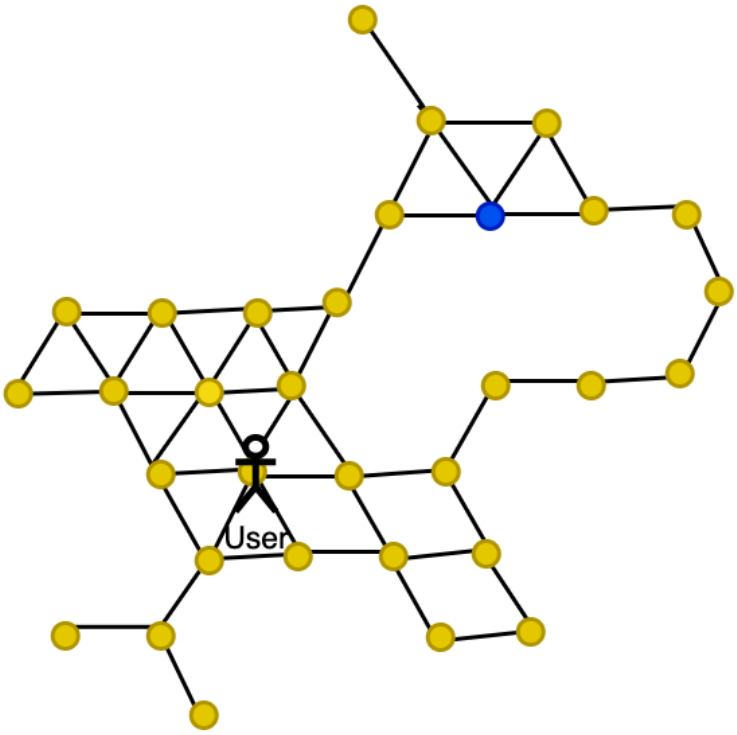

Figure 4.4: Network Structure Graph in The Memory Storage Unit

4.7 Interface design

The NAAD App Inventor Interface utilizes the Unity platform. Unity offers a highly intuitive visual programming interface for various development tasks. Figure 4.5 shows the user interface created with Unity. The user can enter the system's safe or query mode by voice command or clicking the on-screen safe button or query button. When the user enters safe mode, the system alerts the user to the nearest obstacle. When the user enters query mode, as shown in Figure 4.5, The target object is the remote control which the user searches through the voice command. The system automatically detects and marks the target object's category and Position node (blue sphere). Figure 4.1 shows that the system dynamically creates the shortest safe path (blue trajectory) from the user to the target object (remote control) by recording the user's movement Position nodes in real time. These yellow spheres represent the valid Position nodes the user walks indoors. Figure 4.1, from left to right, shows the entire process that the user finds the target object from one room to another according to the navigation path provided

Figure 4.5: Main Menu UI of the NAAD Application System.

by the system. If the user clicks the speaker button on the interface, the system will broadcast the current navigation information to the user. If the user clicks the microphone button on the interface, the system will receive the user's new voice command and proceed with the corresponding operation.

4.8 Results and Discussions

In the experiment, user can choose any of the following ways to use the smartphone equipped with the NAAD assistance system according to their personal preferences: (1) User hold the smartphone by themselves; (2) User insert the smartphone into a portable in the sling pouch and hang it around the neck. Figure 4.6 shows the proposed wearable system. It allows VIB users to free their hands during navigation with the NAAD system. So they can still perceive the surrounding environment according to their daily behaviors, thus reducing the risk of accidents.

The experiment aims to test the localization and navigation capabilities of the NAAD system by verifying its ability to locate a lost object (a remote control) in different spaces. The experiment was conducted in the library of Auburn University, and the experiment location consisted of two rooms. First, we used the NAAD system in the right room to mark the remote

Figure 4.6: Proposed wearable mobility NAAD system by using iPhone 12 Pro Max (right) and sling pouch (left).

as a vulnerable target and placed the target (the remote) on the table. Then we started walking around randomly until we entered the room on the left. Then we will turn on the query mode of the NAAD system in the left room and use voice commands to allow the system to assist us in locating the target item (remote control) and navigating it.

Figure 4.7 (top) shows a yellow sphere representing a user-moved valid safe position node. As shown in the figure, many position nodes form a dynamic position graph. All the security nodes are obtained through the security path that the user walks. The distance between each safety node is set to 0.2 meters, which ensures that no obstacles will be encountered in the navigation path in the subsequent system navigation. As shown in Figure 4.7 (above), there are no safety nodes overlapping with obstacles (yellow chairs) in the dynamic position graph. Figure 4.7 (bottom) shows that after the user moves from the right room to the left room, after turning on the navigation function of the NAAD system, the system uses the dynamic position graph to plan the shortest safe path (blue path) to find the target item in the right room (remote control).

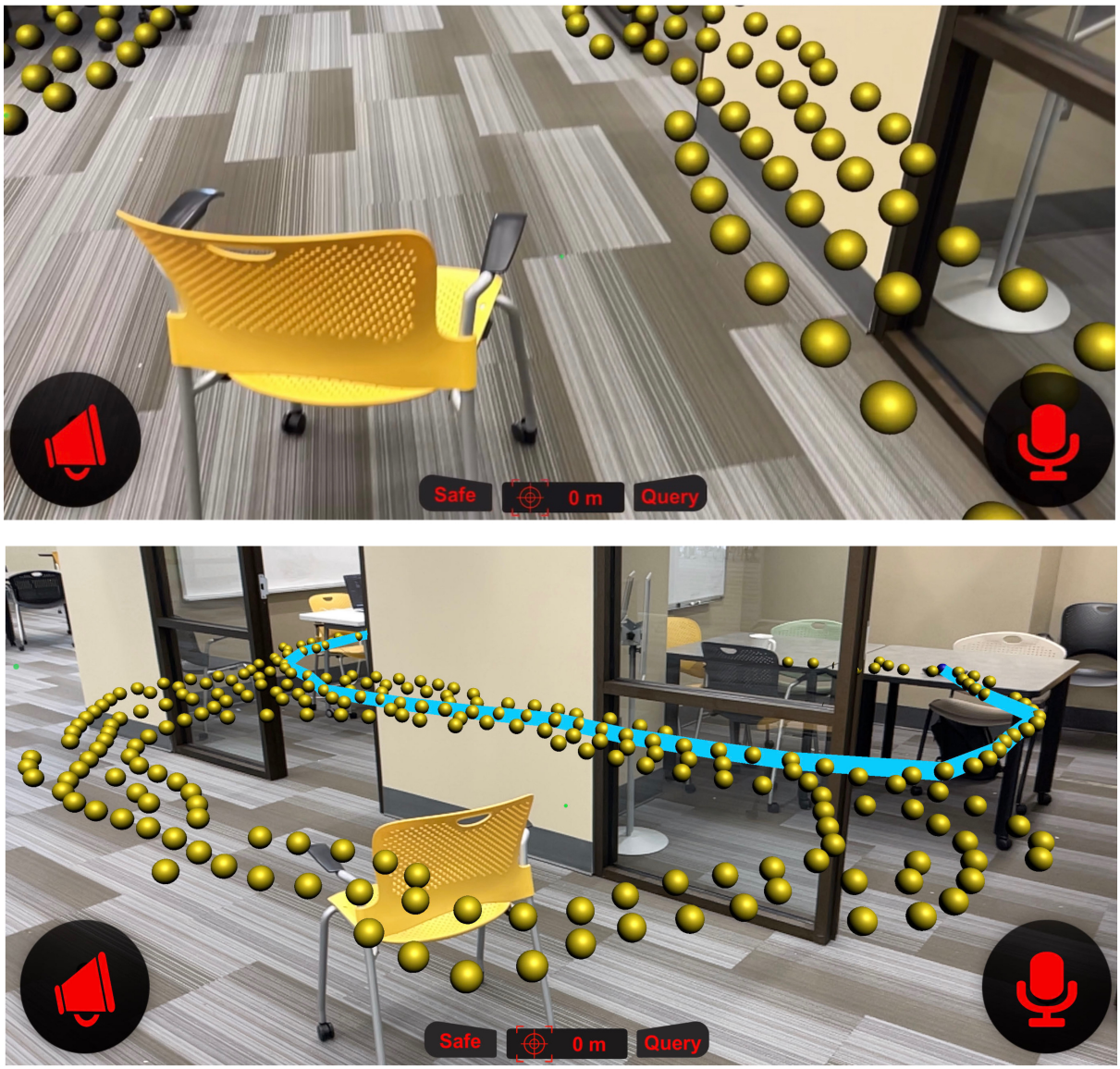

Figure 4.7: The system continuously generates safe yellow location nodes as the user moves (up). The shortest blue navigation path is generated by the system (down).

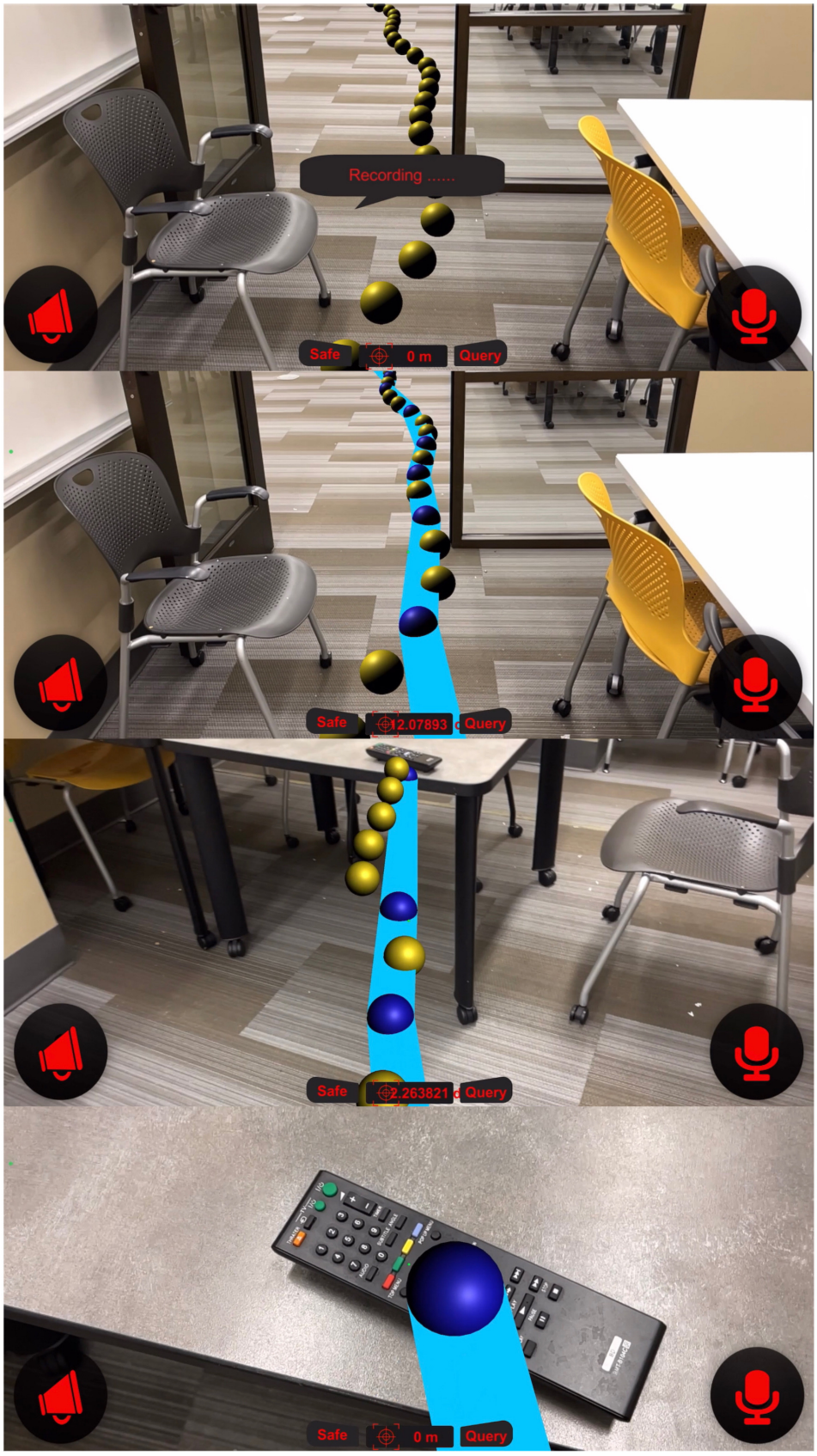

Figure 4.8: The NAAD system helps the user find the specified lost item (remote control)

We also made an Accessible Design for the navigation function module of the NAAD system. Figure 4.8(a) shows that the NAAD system uses the voice module to receive the user's voice commands. When the system receives the user's "search remote" through voice command, the system will immediately activate the memory storage unit and retrieve the location information about the target object. Once the location of the object is found, the system generates the shortest safe path through the dynamic position graph based on the location of the user and the target object. As shown in Figure 4.8(b), the blue path shows the shortest safe path for the user from the left room to the right room. In the distance calculation icon at the bottom of the main interface, the system will display the navigation path's real-time distance (12.08 meters) and broadcast and guide the user along the shortest safe path through the navigation voice system. According to the voice prompts of the navigation system, the user can move left, right, or forward. In our experiments, we set the bias angle to 15 degrees because humans cannot match the orientation of the phone with the navigation direction exactly every time, like a machine. When the angle between the user's direction and the safe path's direction exceeds 15 degrees, the system will voice prompt the user to turn left or right. Otherwise, the system prompts the user to move forward; Figure 4.8(c) shows that the distance from the user to the target object is 2.26 meters. Finally, in Figure 4.8(d), when the user is less than 30cm away from the target object, the system will prompt the user that the target object has been successfully found. At this point, the user can try to reach out and grab the target object. The experiment verifies the positioning and navigation capability of the NAAD system and provides a reference for the practical application of similar problems. In addition, the voice command and navigation system used in the experiment can be widely used in practical applications such as smart homes and smart hospitals to help users locate and navigate target items.

In this experiment, to detect the recognition accuracy of the NAAD system for the target objects grasped and placed by the user at close range, we selected the distance measurement range from 10 cm to 50 cm. Table 4.1 shows the recognition accuracy of the NAAD system for five daily objects at different distances. Each data is obtained by averaging 1000 consecutive data samples. For example, when the mobile phone camera is 30 cm away from the target object, the average recognition accuracy of the system for the target object (cup) is 82.11%.

Target Object	Average Accuracy				
	10 cm	20 cm	30 cm	40 cm	50 cm
Scissors	75.22%	81.93%	82.59%	81.39%	80.26%
Cell phone	65.03%	82.04%	83.17%	82.73%	82.88%
Remote	65.35%	74.14%	81.81%	83.19%	82.10%
Mouse	61.83%	83.203%	82.51%	74.10%	73.99%
Cup	74.48%	82.43%	82.11%	81.94%	78.44%

Table 4.1: Object Detection Accuracy of The NAAD Application System.

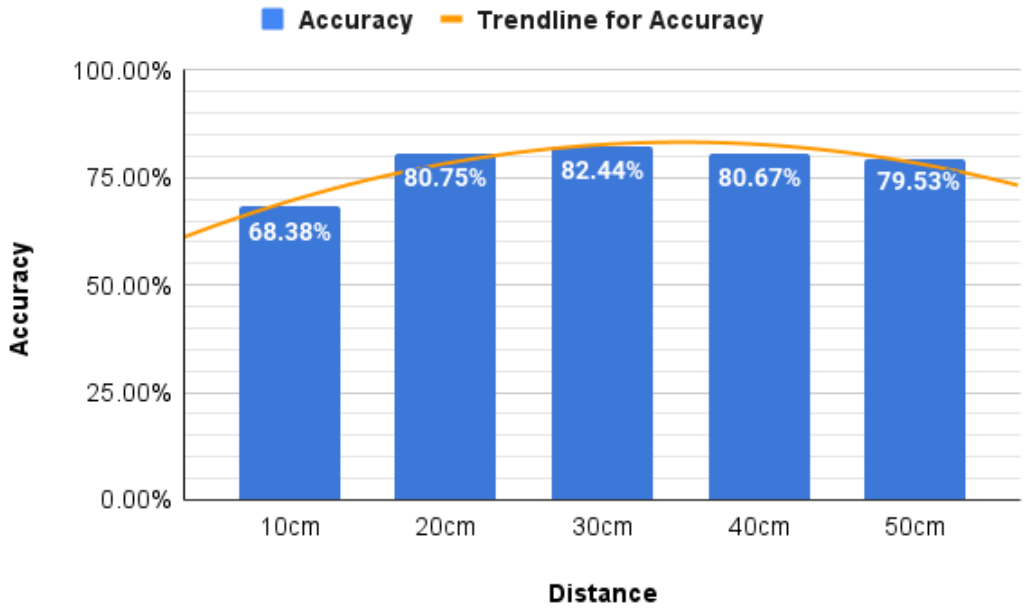

Figure 4.9: Average object detection accuracy at different distances

Figure 4.9 shows the average recognition accuracy curve of the NAAD system for five kinds of daily objects at different distances. We found that the average recognition accuracy of the system peaked when the target object was 30 cm away from the phone camera. When the distance value is shifted from the center of 30 cm to both sides to 10cm and 50 cm, the av erage recognition accuracy of the system gradually decreases. Among them, the recognition accuracy of the system is only 68.38% at the ultra-close distance (d <10 cm). It is because the data for the model training in the NAAD system comes from the COCO dataset, which has relatively few pictures of ultra close-range objects. To solve this problem, we will create a dataset of captured images containing a large amount of ultra close daily items and continuous training

our model in the future. It can effectively improve the item recognition accuracy within this range of the NAAD system [38] [39].

Target Object	Scissors	Cell phone	Remote	Mouse	Cup
Successful found the target object counts (Max of 20)	20	20	20	20	20
Successful Voice navigation prompt counts (Max of 20)	20	20	20	20	20

Table 4.2: Successful Find target objects and navigation voice prompts counts

We conducted 20 navigation tests for each target object using the NAAD navigation system at the Ralph Brown Draughon Library at Auburn University. Table 4.2 indicate that the system achieved a 100% success rate in providing voice prompts for navigation, locating, and guiding to the lost item in different spaces.

4.9 Conclusions

We design and optimize a mobile application called NAAD based on Augmented Reality(AR) Deep Learning, and LiDAR technology, which can assist the visually impaired in finding lost items in different spaces. Among them, the application of AR technology improves user experience, LiDAR technology ensures the accuracy of location information, and Deep Learning technology enables the system to adapt to changing environments and provide the shortest safe path. At the same time, this mobile application eliminates the need to purchase additional assistive devices for the visually impaired. At the same time, this system does not have any network requirements for users. The object navigation module in the NAAD system is a novel, cutting-edge solution to the field of indoor navigation for VIB people. It provides users with target object navigation assistance by recording the target object and the user's location in real-time. It provides an effective and more efficient solution for VIB people to find lost objects, even if the user is far away from the target item or in a different space. Many experimental results show that the NAAD system has high accuracy and security. The system also has a high FPS, which makes it faster and more efficient than comparable systems. In the future, we plan to

add a safety mode to the query mode to improve the safety of users when using the navigation function to find the target object.

Chapter 5

Conclusion and Future Work

We have successfully developed an assistive mobile application - Navigation Assistance through Augmented Reality and Deep Learning (NAAD) system. This system is a revolutionary mobile application designed to solve the difficulties faced by visually impaired and blind people in their daily lives due to a lack of access to visual information. It can be controlled via gestures and a voice command-enabled user interface. This system combines Augmented Reality(AR), Deep learning, and LiDAR technology to provide a comprehensive solution for object detection, user-specified lost object navigation, obstacle detection, and distance calculation. In addition, it also contains an alarm system that analyzes the environment in real-time and reminds the user to avoid obstacles. It eliminates the need to purchase additional assistive devices for the visually impaired. The NAAD system can be installed on any smart mobile device with LiDAR capability, such as iPhone, IPad, etc. In addition, the NAAD system does not require the user to be connected to the Internet, making it an accessible solution for people in areas with no network coverage or poor signal. We found that using Unity's visual programming interface to develop the NAAD system provides an intuitive and user-friendly interface for various development tasks and can support multiple platforms when deploying the application. The NAAD system is designed with a barrier-free user interface. Each function icon of the user interface of the system adopts a large and contrasting click button with red letters on a black background and is equipped with a real-time voice broadcast. These humanized designs have fully considered the particular needs of the visually impaired so that they can more conveniently assist VIB people in their daily lives. The NAAD system is a project that utilizes mobile devices equipped

with LiDAR scanners for precise distance measurement. The system uses LiDAR technology to scan the user's surroundings in real time and construct a virtual 3D world through AR (ARKit) technology. Then use the TensorFlow Lite framework and the Mobile-Net Single-shot Detection deep learning model to enable the system to perform real-time analysis of the user's surroundings through voice commands. The system then localizes the target object by converting the 2D coordinates to 3D coordinates. The system has passed many experiments, and we can conclude that the unique contributions of the NAAD system are summarized as follows:

- We are the first to propose and design a mobile-based indoor navigation system that helps visually impaired people find lost items in different spaces with them. The NAAD system adopts a visual scanning system integrating AR, LiDAR, and Deep Learning to store the location information of users and lost target objects in different spaces and dynamically generates the shortest safe path to provide users with navigation assistance.

- The average Distance Accuracy of the NAAD system in complex environments exceeded 95% within 2 meters and 96% within 5 meters. They were outpacing research paper [23] with an average distance accuracy of 88% within 2 meters.

- The NAAD system ensures the user's safety in navigation based on three aspects: (i) The navigation system creates the shortest safe path based on the Position node of the target object and the Position node generated by the user's movement. It prevents the user from colliding with obstacles such as walls or furniture in the subsequent navigation use. (ii) The real-time voice reminder and vibration function can prevent users from deviating from a safe path during navigation. (iii) The NAAD system adopts Light Detection and Ranging (LiDAR) technology. Its ranging accuracy is high, which can significantly improve the accuracy of the user's position information and the target object.

- The average response time of the NAAD system to recognize objects is 19 milliseconds, which is better than the 100 milliseconds recommended by [44].

- The FPS of the NAAD system is over 30 frames per second and outperforms similar systems [34] and [23].

- The NAAD system's voice broadcast accuracy rate achieved 100%.
- The NAAD system gets a success rate of 100% percent in locating and navigating the lost item in different spaces.

For future work, the design of the NAAD system will be refined incrementally based on feedback from the VIB people. In addition, we plan to work with ophthalmologists and AR experts to gather insights about the NAAD user experience and make necessary improvements to the system to ensure the best user experience. In the future, we plan to integrate a safety mode into the query mode to improve user safety when using the navigation function to find the target object. Finding lost items is a common problem. Many people, especially VIBs, can benefit from our system. It Makes the NAAD application a significant design to improve people's quality of life.

AR and Deep Learning technologies can revolutionize how VIB people find lost objects and navigate their environment to improve their quality of life. We also hope our work will inspire other researchers and designers to continue exploring new and innovative solutions for the visually impaired and other special needs groups.

References

[1] Martinez-Sala, Alejandro Santos, et al. ”Design, implementation and evaluation of an indoor navigation system for visually impaired people.” Sensors 15.12 (2015): 32168-32187.

[2] Nakajima, Madoka, and Shinichiro Haruyama. ”New indoor navigation system for visually impaired people using visible light communication.” EURASIP Journal on Wireless Communications and Networking 2013.1 (2013): 1-10.

[3] Bilgi, Serdar, Ozge Ozturk, and Ayse Giz Gulnerman. ”Navigation system for blind, hearing and visually impaired people in ITU Ayazaga campus.” 2017 international conference on computing networking and informatics (ICCNI). IEEE, 2017.

[4] Shaikh, Farooq, Vishal Kuvar, and Mohd Abbas Meghani. ”Ultrasonic sound based navigation and assistive system for visually impaired with real time location tracking and Panic button.” 2017 2nd International Conference on Communication and Electronics Systems (ICCES). IEEE, 2017.

[5] Botre, Monika Ramchandra, and Anjali R. Askhedkar. ”LiFi and voice based indoor navigation system for visually impaired people.” 2019 IEEE Pune Section International Conference (PuneCon). IEEE, 2019.

[6] Saber, Hakar Mohsin, Nawzad Kameran Al-Salihi, and Rebaz Mohammed Dler Omer. ”Visually Impaired People Navigation System using Sensors and Neural Network.” 2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS). IEEE, 2022.

[7] Kumar, Nitin, and Anuj Jain. ”Smart navigation detection using deep-learning for visually impaired person.” 2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES). IEEE, 2021.

[8] Mahmud, Saifuddin, et al. ”A vision based voice controlled indoor assistant robot for visually impaired people.” 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). IEEE, 2020.

[9] Chaccour, Kabalan, and Georges Badr. ”Computer vision guidance system for indoor navigation of visually impaired people.” 2016 IEEE 8th international conference on intelligent systems (IS). IEEE, 2016.

[10] Barontini, Federica, et al. ”Integrating wearable haptics and obstacle avoidance for the visually impaired in indoor navigation: A user-centered approach.” IEEE transactions on haptics 14.1 (2020): 109-122.

[11] Hutabarat, Dony, et al. ”LiDAR based obstacle avoidance for the autonomous mobile robot.” 2019 12th International Conference on Information Communication Technology and System (ICTS). IEEE, 2019.

[12] King, Fraser, Richard Kelly, and Christopher G. Fletcher. ”Evaluation of LiDAR-derived snow depth estimates from the iPhone 12 pro.” IEEE Geoscience and Remote Sensing Letters 19 (2022): 1-5.

[13] Luetzenburg, Gregor, Aart Kroon, and Anders A. Bjørk. ”Evaluation of the Apple iPhone 12 Pro LiDAR for an application in geosciences.” Scientific reports 11.1 (2021): 22221.

[14] O’Keeffe, Rosemary, et al. ”Long range LiDAR characterisation for obstacle detection for use by the visually impaired and blind.” 2018 IEEE 68th Electronic Components and Technology Conference (ECTC). IEEE, 2018.

[15] Chitra, P., et al. ”Voice Navigation Based guiding Device for Visually Impaired People.” 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). IEEE, 2021.

[16] Galatas, Georgios, et al. ”eyeDog: an assistive-guide robot for the visually impaired.” Proceedings of the 4th international conference on pervasive technologies related to assistive environments. 2011.

[17] Zhang, Yingzhi, et al. ”Perception framework through real-time semantic segmentation and scene recognition on a wearable system for the visually impaired.” 2021 IEEE International Conference on Real-time Computing and Robotics (RCAR). IEEE, 2021.

[18] Ton, Carolyn, et al. ”LiDAR assist spatial sensing for the visually impaired and performance analysis.” IEEE Transactions on Neural Systems and Rehabilitation Engineering 26.9 (2018): 1727-1734.

[19] Busaeed, Sahar, et al. ”LidSonic for Visually Impaired: Green Machine Learning-Based Assistive Smart Glasses with Smart App and Arduino.” Electronics 11.7 (2022): 1076.

[20] Kalpana, S., et al. ”Voice recognition based multi robot for blind people using LiDAR sensor.” 2020 International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2020.

[21] Dragne, Ciprian, et al. ”Distance Assessment by Object Detection—For Visually Impaired Assistive Mechatronic System.” Applied Sciences 12.13 (2022): 6342.

[22] Huang, Jonathan, et al. ”An augmented reality sign-reading assistant for users with reduced vision.” PloS one 14.1 (2019): e0210630.

[23] Chen, Hsuan, et al. ”The obstacles detection for outdoor robot based on computer vision in deep learning.” 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin). IEEE, 2019.

[24] Lin, Bor-Shing, Cheng-Che Lee, and Pei-Ying Chiang. ”Simple smartphone-based guiding system for visually impaired people.” Sensors 17.6 (2017): 1371.

[25] Zhu, Jing, and Yi Fang. ”Learning object-specific distance from a monocular image.” Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[26] Hua, Minjie, Yibing Nan, and Shiguo Lian. ”Small obstacle avoidance based on RGB-D semantic segmentation.” Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.

[27] Jiang, Zhenghan, Qiangfu Zhao, and Yoichi Tomioka. ”Depth image-based obstacle avoidance for an in-door patrol robot.” 2019 International Conference on Machine Learning and Cybernetics (ICMLC). IEEE, 2019.

[28] Kang, HyeongYeop, Geonsun Lee, and JungHyun Han. ”Obstacle detection and alert system for smartphone ar users.” Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology. 2019.

[29] Troncoso Aldas, Nelson Daniel, et al. ”AIGuide: An augmented reality hand guidance application for people with visual impairments.” Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility. 2020.

[30] ARAnchor. Retrieved June 15, 2021, from https://developer.apple.com/documentation/arkit/aranchor

[31] Dhou, Salam, et al. ”An IoT machine learning-based mobile sensors unit for visually impaired people.” Sensors 22.14 (2022): 5202.

[32] Ou, Soobin, Huijin Park, and Jongwoo Lee. ”Implementation of an obstacle recognition system for the blind.” applied sciences 10.1 (2019): 282.

[33] Lin, Yimin, et al. ”Deep learning based wearable assistive system for visually impaired people.” Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.

[34] Srinivasan, Akshaya Kesarimangalam, Shwetha Sridharan, and Rajeswari Sridhar. ”Object localization and navigation assistant for the visually challenged.” 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.

[35] Xie, Tianshi, and Cheryl D. Seals. ”Mobile augmented reality using deep learning for visually impaired people.” ACM SIGACCESS Accessibility and Computing 132 (2022): 1-1.

[36] Permana, Silvester Handy, et al. ”Comparative analysis of pathfinding algorithms a*, dijkstra, and bfs on maze runner game.” IJISTECH (International J. Inf. Syst. Technol., vol. 1, no. 2, p. 1 (2018).

[37] Chen, Jolly, and Robert G. Belleman. ”MeasVRe: Measurement Tools for Unity VR Applications.” 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2022

[38] Tong, Kang, Yiquan Wu, and Fei Zhou. ”Recent advances in small object detection based on deep learning: A review.” Image and Vision Computing 97 (2020): 103910.

[39] Pundir, Arun Singh, and Balasubramanian Raman. ”Dual deep learning model for image based smoke detection.” Fire technology 55.6 (2019): 2419-2442.

[40] Understanding Success Criterion 2.5.5: Target Size. Retrieved June 04, 2021, from https://www.w3.org/WAI/WCAG21/Understanding/target-size.html

[41] Apple Human Interface Guidelines: Adaptivity and Layout: Target Size. Retrieved June 04, 2021, from https://developer.apple.com/design/human-interface-guidelines/foundations/layout/

[42] Material Design Accessibility. Retrieved June 04, 2021, from https://m2.material.io/design/usability/accessibility.html#understanding-accessibility

[43] Bodine, Cathy. Assistive technology and science. SAGE Publications, 2012.

[44] Kaaresoja, Topi, Stephen Brewster, and Vuokko Lantz. ”Towards the temporally perfect virtual button: touch-feedback simultaneity and perceived quality in mobile touchscreen press interactions.” ACM Transactions on Applied Perception (TAP) 11.2 (2014): 1-25.

[45] Troncoso Aldas, Nelson Daniel, et al. ”AIGuide: An augmented reality hand guidance application for people with visual impairments.” Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility. 2020.

[46] ARAnchor. Retrieved June 15, 2021, from https://developer.apple.com/documentation/arkit/aranchor%20.