DOCTORAL THESIS

DEVELOPMENT OF A WEARABLE ASSISTIVE DEVICE FOR NAVIGATION FOR THE VISUALLY IMPAIRED WITH COMMAND AND REQUEST SUPPORT.

Author:
Bismark Kweku Asiedu Asante

Supervisor:
Professor Hiroki Imamura

A thesis submitted in fulfillment of the requirements
for the Doctor of Philosophy

of the

Faculty of Science and Engineeering
Graduate School of Engineering

February 19, 2024

Declaration of Authorship

I, Bismark Kweku Asiedu Asante, declare that this thesis titled, “DEVELOPMENT OF A WEARABLE ASSISTIVE DEVICE FOR NAVIGATION FOR THE VISUALLY IMPAIRED WITH COMMAND AND REQUEST SUPPORT.” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at Soka University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________________

Date: ______________________________

Acknowledgements

I would like to express my heartfelt gratitude to the many individuals and institutions that have supported me throughout my doctoral journey. This dissertation would not have been possible without their unwavering encouragement, guidance, and contributions.

First and foremost, I am deeply thankful to my dissertation advisor, Prof. Hiroki Imamura, for their invaluable mentorship, unwavering support, and insightful feedback. Their expertise and dedication have been instrumental in shaping this research. I am grateful for the countless hours of guidance and encouragement they provided, which significantly enriched the quality of this work.

I would like to extend my appreciation to the members of my dissertation committee, Prof. Hidehiko Shishido and Prof. Dr. Tatsuo Unemi, for their valuable insights and constructive criticism that helped refine my research. Their diverse perspectives and expertise contributed significantly to the academic rigor of this dissertation.

I am grateful to my colleagues and peers who offered their camaraderie and shared intellectual discourse. The following are worthy of mention for their support; Dr. Clifford Broni-Bediako, Mr Hideaki Saito, Mr. Diego Alberto Herrera Ollachica and Mr. Ryu Nomura. Their feedback, discussions, and collaboration have been instrumental in shaping my ideas and refining my research.

I would also like to acknowledge the financial support provided by Makiguchi Foundation Scholarship. Their support allowed me to dedicate my time and energy to this research without the added burden of financial constraints.

My deepest gratitude goes to my family for their unwavering love, encouragement, and understanding. Their belief in my abilities and their sacrifices have been my constant source of strength.

Last but not least, I extend my appreciation to all the participants and volunteers who participated in this research. Their willingness to contribute their time and insights was essential to the success of this study.

In closing, I am thankful for the rich tapestry of experiences, knowledge, and support that has enriched my doctoral journey. This dissertation stands as a testament to the collaborative efforts of many, and I am deeply appreciative of everyone who played a part in its realization.

SOKA UNIVERSITY

Abstract

Faculty of Science and Engineeering
Graduate School of Engineering

Doctor of Philosophy

DEVELOPMENT OF A WEARABLE ASSISTIVE DEVICE FOR NAVIGATION FOR THE VISUALLY IMPAIRED WITH COMMAND AND REQUEST SUPPORT.

by Bismark Kweku Asiedu Asante

Vision loss is the most severe sensory disability that renders a person nearly immobile, with the fear of bumping into obstacles or becoming lost in the environment. The risks associated with blindness go beyond inconvenience and can lead to falls and injuries. Therefore, care for the blind requires excellent guidance. To safely guide and navigate the environment, wearable assistive devices for guiding the visually impaired require a crucial component which is obstacle avoidance. Obstacle avoidance refers to the ability of a system or an individual to detect, recognize, and navigate around obstacles to avoid collisions or disruptions in movement

We propose the design and implementation of a wearable assistive navigation device, WAND, that integrates a novel obstacle avoidance strategy with other functional components for navigation, ground plane checking, and request assistance. The system utilizes an advanced sensing camera, a stereo camera, ZED2 to stream images of the environment, and real-time processing capabilities provided by a microcontroller, Jetson Nano, to detect and analyze the environment with a custom system based on the YOLOv5 framework. By prioritizing locations, we aim to identify areas that are safe for navigation from hazardous obstacles and confusing landmarks such as entrances, intersections, and potential hazards. Simultaneously, the system focuses on detecting obstacles in close proximity to the user, monitoring the flatness of the ground, ensuring the user is on the right track, and providing immediate feedback and guidance to avoid collisions or potential dangers.

To evaluate the performance and effectiveness of our proposed wearable assistive device, we conducted comprehensive field tests in both indoor and outdoor environments. These experiments simulated real-world scenarios, allowing us to assess the system's reliability, accuracy, and user experience. Our evaluation criteria included response time, obstacle detection rate, accuracy of the detection, and prioritizing the most hazardous obstacle to alert the visually impaired user.

The results obtained from our evaluations demonstrate the efficacy and robustness of the proposed obstacle avoidance strategy. The system effectively provides visually impaired individuals with timely and relevant information, enabling them to navigate safely and independently in various environments. By addressing the unique challenges faced by VIPs, our research contributes to improving their quality of life and fostering their inclusion in society

Vision Impairment, Electronic Travel Aids, Wearable Device

Contents

List of Figures

List of Tables

List of Abbreviations

AMD	**A**ged-Related **M**acular **D**egeneration
ASR	**A**utomated **S**peech **R**ecognition
CBAK	**C**omposite MOS Predictor for **B**ackground-Noise Intrusiveness
CNN	**C**onvolutional **N**eural **N**etwork
COVL	**C**omposite MOS Predictor of **O**verall Signal
CSIG	**C**omposite Measure of **S**ignal-to-Distortion Ratio
CPD	**C**ycles **P**er **D**egree
DNN	**D**eep **N**eural **N**etworks
DSEGAN	**D**eep **S**peech **E**nhancement **G**enerative **A**dversarial **N**etwork
DPM	**D**eformable **P**arts **M**odels
ETA	**E**lectronic **T**ravel **A**ids
EMC	**E**lectronic **M**obility **C**ane
EOA	**E**lectronic **O**rientation **A**ids
FP	**F**alse **P**ositives
FN	**F**alse **N**egatives
GAN	**G**enerative **A**dversarial **N**etwork
GPS	**G**lobal **P**osition **S**ystem
HMM	**H**idden **M**arkov **M**odels
HOG	**H**istogram **O**riented **G**radients
ICD	**I**nternational **C**lassification **D**iseases
ISEGAN	**I**terated **S**peech **E**nhancement **G**enerative **A**dversarial **N**etwork
MFCC	**M**el- **F**requency **C**epstral **C**oefficients
PESQ	**P**erceptual **E**valuation of Speech **Q**uality
PLD	**P**osition **L**ocator **D**evices
RFID	**R**adio **F**requency **ID**entification
R-CNN	**R**egional-Proposal **C**onvolutional **N**eural **N**etwork
SASEGAN	**S**elf-**A**ttention **S**peech **E**nhancement **G**enerative **A**dversarial **N**etwork
SEGAN	Speech Enhancement Generative Adversarial Networks
SDG	**S**ustainable **D**evelopment **G**oal
SSD	**S**ingle **S**hot-Model **D**etector
SSNR	**S**egmental Signal-to-**N**oise **R**atio
STOI	**S**hort-**T**ime Objective **I**ntelligibility
TTS	**T**ext **T**o **S**peech
TP	**T**rue **P**ositives
TP	**T**rue **N**egatives
VIP	**V**isually **I**mpaired **P**erson
WAND	**W**earable **A**ssitive **N**avigation **D**evice
YOLOv5	**Y**ou **O**nly **L**ook **O**nce version 5

Chapter 1

Introduction

1.1 Background

Visual impairment, also known as vision impairment or visual disability, refers to a range of conditions where a person's ability to see is significantly reduced or impaired to the extent that it impacts their daily life, activities, and functioning. Visual impairment can vary in severity and can be caused by various eye conditions or diseases [1]. The specific criteria for classifying visual impairment may vary by region and medical standards, but it generally encompasses the following categories: People with low vision have significant visual limitations that cannot be fully corrected with standard eyeglasses, contact lenses, or medical treatment. They may still have some residual vision, which can be useful for various tasks, but they often require visual aids and adaptations to perform daily activities. Low vision can range from mild to severe. Legal blindness is a specific level of visual impairment defined by government regulations or healthcare authorities for various purposes, such as eligibility for disability benefits, services, or special considerations [2]. The criteria for legal blindness often include one or more of the following: Best corrected visual acuity (with glasses or contacts) in the better eye of 20/200 for the United States of America or 20/400 for World Health Organization [3]. Visual field limitation to 20 degrees or less in the better eye. Total blindness, also known as complete blindness, refers to the complete absence of vision or light perception. Individuals with total blindness rely on non-visual cues and assistive devices, such as braille, guide dogs, and canes, to navigate their surroundings and perform daily tasks. The eye like the other organs of the human body can sometimes fail in performing its role due to abnormalities in some of the components of the eye. Often this is due to aging, defects caused by injury, or some hereditary defects. Some of the well-known diseases of the eye include refractive Error, Glaucoma, Cataracts, Diabetic retinopathy, and Aged-Related Macular Degeneration (AMD) [4]. These eye diseases may result in blurred vision or total loss of sight or vision. For instance, AMD could make the patient have dark patches in the images formed by the defective eye therefore the brain cannot recognize the objects located in the dark spots.

According to World Health Organization, [5], there are around 2.2 billion with eye problems, near or distance vision impairment, and half with preventable visual impairments. The number of visually impaired persons (VIP) is estimated to grow and this growth indicates that visual impairment is one of the major health issues facing the world [6], [7]. Even though early vision screening for some eye diseases such as refractive error may decrease the risk or prevent a great proportion of individuals from vision loss or blindness, the cost and time involved often deter them from seeking medical attention. R Varma *et. al* [8] indicated that in the United States of America, the risks of vision impairment among the populations increase as they age between 40 years to 80 years old.

People with severe visual impairment face numerous challenges. These challenges can significantly impact their ability to move safely and independently in both familiar and unfamiliar environments [9]. Adele Crudden [10] indicated that using public transportation can be difficult, as blind individuals may struggle to locate bus stops, read schedules, or ensure they board the correct vehicle. Accessible public transportation services are essential to their mobility. These challenges encompass difficulties in mobility, understanding and navigating their surroundings, and independently performing daily activities.

VIPs may struggle to detect and navigate around obstacles in their path. They often rely on physical landmarks as cues to help locate and verify their locations [11]. Sometimes these physical landmarks become obstacles or impediments in the navigation paths. Obstacles include common items in both indoor and outdoor environments like furniture, stairs, curbs, and even unexpected obstacles such as construction sites or fallen objects. In some situations, pedestrians do become obstacles in the path of the VIP too. These hurdles arise because many activities and actions are contingent on visual perception and understanding of the environment. Navigational difficulties confine visually impaired individuals (VIPs) to familiar areas with fewer obstacles, adversely impacting their economic and social lives [12], [13]. Several studies have been carried out to assess the effect and the impact that vision impairment has on the individual, family, and their communities. Juodvzbaliene *et. al* [14]'s studies show that the subject tends to rely on close relatives or friends for support in performing daily activities. The reliance often results in the restriction of participation in daily activities [9].

In most parts of the world, VIPs are faced with various challenges such as mobility if they are not assisted. They will bump into obstacles and possibly risk being injured. In order to prevent these hazards, there are constant efforts aimed at assisting the VIPs to be independent and navigate their path safely. Also, the Sustainable Development Goals (SDGs) consider the promotion of eye health and assistive technology for navigation as its agenda as the loss of sight leaves 90% of the VIPs in low-income and middle-income countries with to be left behind and have no source of income.

In light of this, many devices and tools are being developed aimed at assisting the VIPs to move freely and engage in their daily activities. Some of these are braille for reading, a guide dog for walking around, obstacle avoidance, and assistive devices for reading or walking around. Most recently, wearable assistive equipment has been given some attention as a better solution for helping the blind with their mobility and daily tasks [15].

Even though, the emergent of these devices, there are still a lot of issues and challenges with devices being developed hence the lack of interest in the use of the device by the visually impaired.

1.2 Problem Statement

In contemporary society, an individual typically dedicates approximately 1.1 hours each day to commuting and allocate a consistent portion of their income to cover transportation expenses. Furthermore, it is anticipated that the duration of daily commuting will notably rise by the year 2050 [16]. This necessity of commuting raises more concerns for the VIP who faces a lot of challenges with mobility. People with visual impairments face significant challenges in navigating their surroundings independently and safely. Even though there are assistive devices for the VIP to use in navigating, such as guide dogs, white canes are categorized as traditional navigation aids. The use of traditional navigation aids, such as canes and guide dogs for aid mobility, has limitations in providing real-time and context-aware information about the environment [17]. It is costly for the blind to have nursing home care. From our review, assistive devices are cheaper, and a well-developed one can be a more convenient assistant to the user. Using assistive devices, the problems with their mobility are very complicated and challenging. Mobility involves knowing the immediate environment, orientation, and obstacle avoidance. Most of the assistive devices available now have good results with obstacle avoidance and perform woefully with giving information on the immediate environment and orientation of the user. To address this issue, there is a critical need for the development of wearable assistive navigation systems that can enhance the mobility, autonomy, and safety of individuals with visual impairments. These systems must overcome various technical, usability, and accessibility challenges to provide seamless and reliable assistance, ensuring that users can navigate diverse environments with confidence and ease.

Even though there are electronic travel aids out there, which perform well technologically but are yet to be taken up by potential users, that is VIP? Does the VIP not have challenges with navigation or traveling? Do they not want devices to assist their navigation? Why is it that these technologically successful electronic travel aids are not being used by VIPs?

Unfortunately, however, none have matured into widely used tools or have gained acceptance in the blind community because of the many problems involved in using them in real-world use cases such as their size, weight, battery life, reliability, ease of use, cost, interference with other senses and primarily the time required to master them even on a basic level, which is the parameter we focus on here.

1.2.1 Statement of the Problem

The present study work is defined by the following statement of the problem: To develop a wearable assistive device for the VIP to be able to navigate safely between indoor and outdoor environments with full awareness of objects around them and their motions as well.

The objective of the Research The objectives of the research are further divided into the following sub-problems to be addressed:

1. To design and implement a wearable assistive navigation device (WAND) for safely guiding the visually impaired to their desired destination. The functionality of the device should include obstacle avoidance, pathfinding, ground plane monitoring, and command or request support.

2. Provide more accurate information about the obstacles in terms of the distances and locations

3. To formulate a robust obstacle avoidance strategy to guide the system used to avoid bumping into the obstacles.

4. Alert the user about the discontinuities in the plane or ground being traversed by the users.

1.3 Proposed Contribution

Our approach is to develop a system that will guide the user in walking with features that can inform the user through the speech of possible obstacles in their path, the distance to be covered before bumping into the objects, or the number of steps before bumping into the obstacle. To achieve this objective, we developed a visual system to plan the path of the user using a stereo camera, we extracted the floor plan and determined the possible obstacles. Using the depth data captured by the stereo camera we can estimate the horizontal planes and the vertical planes in the image and determine the path with the position of various obstacles and their distance from the camera. This information is important for guiding visually impaired users in their environment.

The recommendations provided in the ETA guidelines [18] suggest that the information vital to assisting the VIP includes: obstacle detection and identifications, path surface discontinuities (thus changes in the floor plane that the user is traversing), shoreline information, and landmark identification. Considering this suggestion as to the needed information by the VIP, we reviewed previous to propose a new novel way of avoiding objects and being aware of the environment.

We would develop a wearable assistive device that can help the visually impaired to have greater freedom and independence in moving around their immediate environments. The device can allow them to request help by issuing commands for assistance. If the project is completed, the device can allow the blind to move around freely and provide them with information about their immediate environment, so they do not bump into obstacles. This can give them opportunities to engage in social activities that would earn them income for their livelihood.

The device would be head-mounted. The components have a microphone and speaker for communication between the user and the device. The device will assist with the information that is given to the user. This research is supported to help the visually impaired make a request for assistance anytime they need it and such support will give them about 90% independence in mobility and their daily life activities. The speech communication model that determines the language in which a device operates can be altered by training it in other languages thereby making it relevant internationally. There are blind people across the world and research is being conducted globally in that regard. Therefore, our research also focuses on meeting the standards set down in developing assistive devices for the blind

1.4 Thesis Organization

For the subsequent chapters of the thesis, we have structured them as follows: Chapter two delves into the relevant literature pertaining to our work, highlighting its unique features. Chapter three offers a comprehensive and intricate account of the wearable assistive navigation system. It covers the system's design, development, implementation, and operation. Chapter four provides an in-depth description of the components of the assistive device, tailored for navigating in both indoor and outdoor environments. Chapter five focuses on the speech communication interface, which users utilize to interact with the device and access its request functionality. We conclude with Chapter Six, offering a summary of our findings and outlining potential avenues for future improvements.

1.5 Conclusion

The chapter provides an introduction and background information on the research work. We describe the problem statement with our research questions and state the contribution of this work to the development of wearable assistive technologies We provide a perspective on the problem we aim to provide a solution. We described the problem and stated the research problem we intended to tackle. We identify that obstacle avoidance is important as well as the VIP being aware of their environment. The proposed solution and how the problems were solved using the proposed solution were presented.

Chapter 2

Literature Review

2.1 Introduction

This chapter is dedicated to reviewing a diverse range of literature pertinent to our research study. This body of literature encompasses technologies that share similarities with our work, studies addressing the specific problem we aim to solve, investigations into analogous projects or initiatives, and research focused on aiding VIPs. The purpose of this comprehensive literature review is twofold: firstly, to cultivate a profound comprehension of the problem at hand, and secondly, to lay the groundwork for delivering an excellent solution to this problem.

2.2 Vision, Vision Impairment, Causes and Effects

2.2.1 Vision and Vision Impairment

The ability to interpret or understand the structures and various items observed in an environment is termed vision or visual perception. There are different types of vision based on the environment; photopic vision is often related to daytime vision, whereas scotopic vision refers to night vision. Vision-related to color is called color vision. Under luminous conditions in humans, photopic vision, mediated by cones, prevails in the light-capture process, with a maximum absorption peak at approximately 555 nm [19].

To determine the visual system's ability to distinguish points and objects in space, a quantitative measure known as visual acuity is employed. This measure is expressed as cycles per degree (CPD), representing angular resolution or how effectively an eye can differentiate one object from another concerning visual angles. The most commonly performed measure is distance visual acuity, involving the placement of a visual chart at a specified distance [20]. Visual acuity serves as a fundamental visual screening test, often used by licensing agencies to assess an individual's fitness for driving. However, it's worth noting that visual acuity requirements vary significantly from state to state in the United States. For instance, in Florida, drivers must have a visual acuity of 20/70 in either eye, with or without corrective lenses, whereas in Connecticut, drivers are required to have 20/40 in the better eye, with or without corrective lenses [21]. In some states, drivers who fail to meet the vision requirement may face restrictions.

Vision impairment, a term used in the medical field, encompasses various forms of vision loss [22]. It serves as a measure of visual acuity in humans, covering both total vision loss and partial impairment. Vision loss can result from a range of factors, including injuries, although complete loss of sight at an early age is relatively rare. Some conditions or groups of diseases leading to severe vision impairment or blindness in children under the age of sixteen (16) are categorized as congenital blindness [23]. The widely recognized standard for classifying vision impairment is the International Classification of Diseases, specifically ICD-11, and this classification is detailed in Table 2.1.

TABLE 2.1: The classification of Vision Impairment and the measured visual acuity as stipulated by the ICD-11

Categories	Description	Measure of Acuity
Category 0	No or Mild Visual Impairment	Having a visual acuity of better than 6/18
Category 1	Moderate Visual Impairment	Having a visual acuity worse than 6/18 and better than 6/60
Category 2	Severe Visual Impairment	Having a visual acuity worse than 6/60 and better than 3/60
Category 3	Blindness	Having a visual acuity worse than 3/60 and better than 1/60
Category 4	Blindness	Having a visual acuity worse than 1/60 with light perception
Category 5	Blindness	Having a visual acuity worse than 1/60 with light perception

2.2.2 Causes of Vision Impairment

Various factors lead to one losing sight. Loss of sight may result from lifestyle exposures and behaviors, genetics, aging, or developing health conditions. Some factors that cause sight loss are aging, disorders due to inherited genes or diet, and physical injuries.

Glaucoma is the most common cause of irreversible blindness worldwide [24]. It occurs as a result of the loss of retinal ganglion cells (RGCs). When detected early, treatment can be offered to prevent loss of sight. Since it is asymptotic it is difficult for patients to notice that they have glaucoma [25] [26]. Cataracts are the clouding of the eye's natural lens, leading to blurred or cloudy vision. They often develop with age but can also be caused by injury, medications, or underlying medical conditions. AMD is a degenerative eye disease that affects the macula, the central part of the retina. It can lead to a loss of central vision, making it difficult to see fine details or read. Diabetic retinopathy is a complication of diabetes that damages the blood vessels in the retina. It can cause vision problems, including blurred vision, floaters, and, in severe cases, blindness. Traumatic eye injuries can result in vision impairment. These injuries can range from minor scratches to severe damage, including retinal tears or detachment.

Early detection and appropriate medical intervention can often prevent or manage many causes of vision impairment. Regular eye exams with an optometrist or ophthalmologist are essential for maintaining good eye health and detecting issues early. Additionally, lifestyle factors, such as a healthy diet and protective eyewear, can help reduce the risk of some vision-related problems

2.2.3 Challenges of Visual Impairments

It is estimated that eighty-five percent of information related to perception, learning, cognition, and activities is through vision [27]. This indicates the importance of eyesight to humans and animals alike. Unfortunately, one has to lose sight partially or completely. Not being able to carry out these activities with ease and independently comes with a lot of challenges.

An inspiring talk at TEDxGouda by Karthik Kannan [28] shed light on the challenges faced by the VIP. He made a clear presentation on how challenging it is as technology advances to make life easier for people but not factoring in the actual needs of the VIP. Imagine how a VIP could use a touch screen or read texts from such devices. Independence was one of the biggest challenges the disabled faced as most of their challenges were not considered in the design of most systems. For most vision, impaired persons, the effect of this condition on their life could be attributed to the activity limitations and access to socioeconomics resources [13][12]

In the absence of an assistive guide, partially sighted or VIP always try to memorize all locations they have been to to become familiar with them. In a new, unfamiliar setting they become desolate due to the difficulty of identifying landmarks and places. This situations often lead to some VIPs being completely dependent on other family and friends to reach their desired destination [29]. The task of route planning in an unforeseen obstacle environment can severely impede the independence and mobility of VIPs and thus reduce their willingness to travel [30].

2.3 Review of Assistive Technologies for the Visually Impaired

In this section, we will review research works on assistive technologies focused on helping the visually impaired. Assistive Technologies refer to any technological innovations (products, environmental modifications, services, and processes) helpful to surmount any challenges for an individual especially persons with disabilities or special needs. There are two contexts of assistive devices: medical applications and social applications. The assistive technologies for the visually impaired fall under the social category. Recently several assistive have been developed to support those with different disabilities as well as people with diseases that make them unable to undertake certain tasks.

2.3.1 Categorization of Mobility Devices

The development of an assistive device for the Blind has been around since 1960. Over the years, several devices have been developed to assist the VIP with navigation. The devices have been categorized by several research works based on the technology and sensors used in the processing of information and the mode of transmitting the outputs [31]–[34]. The devices are classified into three groups namely Vision enhancement, Vision replacement, and vision substitution [31]. Vision enhancement often involves capturing images from the camera, processing the data, and displaying the output on a visual display. It is mostly head-mounted and most commonly seen in visual reality systems. Vision replacement also captures images of scenes, processes them, and displays the information directly to the visual cortex in the brain of humans through the optic nerve [35]. Vision substitution captures the images or video streams, processes the data and the output is non-visual. The most common mediums of transmitting the output in visual substitution are tactual or auditory. Sometimes the non-visual output could be a combination of them. Our research is in the vision substitution category and this category is also divided into another subcategory. Electronic travel Aids, Electronic Orientation aids, and Position Locator devices. These subcategories focus on certain characteristics and functions

2.3.1.1 Electronic Travel Aids, ETAs

Numerous mobility aid devices, called electric travel aids (ETA) [1], were introduced to promote locomotion for the blind. To ensure the safety of locomotion, ETAs have incorporated functions for obtaining information on orientation. For example, an

ETA sensor determines the user's location, the direction the user takes, and the distance of nearby objects. The aids transform visual information about the environment into a form that can be conveyed through another sensory modality. The most common sensory modes are auditory and touch. A large number of ETAs for the blind have been developed over the past 40 years. ETA's are designed to transform environmental information into a form that can be conveyed through other sensory modalities (auditory or tactile) [36]. Electronic Travel Aids are devices designed to assist blind or visually impaired individuals in safely navigating their surroundings and avoiding obstacles. They typically use various sensors and technologies to detect objects, obstacles, or changes in the environment and provide auditory or tactile feedback to the user. Common examples include:

Ultrasonic Canes: These canes emit high-frequency sound waves that bounce off objects in the environment. By analyzing the reflected sound, the user can detect obstacles and determine their proximity.

Laser Canes: Similar to ultrasonic canes, laser canes use laser beams to detect obstacles and provide feedback to the user.

Smart Glasses: Some wearable smart glasses incorporate obstacle detection and recognition features. Cameras and sensors on the glasses capture the environment, and the device provides audio or tactile cues to help the user navigate.

2.3.1.2 Electronic Orientation Aids, EOAs

Electronic Orientation Aids focuses on providing information about the user's orientation and location in their environment. They help users understand their surroundings and maintain a sense of direction. EOAs often utilize GPS technology and other sensors. Key examples include: These devices use GPS technology to provide information about the user's location, nearby points of interest, and directions to a desired destination through audio or tactile feedback. Some EOAs incorporate tactile maps or diagrams that users can explore with their fingers. These tactile representations of the environment help users understand layouts and landmarks. Many smartphone apps are designed to serve as electronic orientation aids. They provide spoken directions, and location information, and can work in conjunction with GPS and digital maps.

2.3.1.3 Position Locator Devices, PLDs

Position Locator Devices are designed to help users find specific objects or locate destinations within an environment. They are often used in conjunction with indoor navigation systems. These devices use technologies like radio frequency identification (RFID) or Bluetooth beacons to pinpoint the location of objects or areas. Examples include: These systems deploy beacons or RFID tags within buildings or indoor spaces. Users with PLDs can receive information about their proximity to these tags and use it to navigate indoor environments accurately. Some PLDs are designed to help users locate specific objects, such as keys, wallets, or personal items, by attaching RFID tags or Bluetooth trackers to these objects and using a handheld or mobile device to locate them.

[31] proposes structural and operational features for researchers, developers, and engineers to include in Electronic Travel Aids. The features were based on discussions with groups of VIPs, software developers, and engineers. These features were used to perform maturity analyses for ETA systems presented in their survey paper. Table 2.2 shows the features and their descriptions.

TABLE 2.2: Structural and operational features

Features	Description
Realtime	The system operates fast enough to provide users with information to to respond to situations
Wearable	The device is worn on the body
Portable	Easy to carry around
Simple	Easy to use or operate
Reliable	Information provided is consistently accurate and quality of performance is good
Robust	Adapt to changes in environmental conditions

2.4 Techniques and Methods Commonly Used in Vision Substituted Systems

2.4.1 Distance estimation and measuring sensors

Knowing the distance of an object or obstacle is an important task in ETA systems. The estimated distance of objects in the proximity of the VIP is often provided as alerts for them to evaluate how to maneuver around such objects. Distance estimation and measuring sensors are devices that use various technologies to determine the distance between the sensor and a target object or surface. These sensors have a wide range of applications, from industrial automation and robotics to automotive safety systems and consumer electronics. Here are some common types of distance estimation and measuring sensors used in ETA systems.

Ultrasonic sensors work based on the transmission and reception of high-frequency sound waves (ultrasonic waves). The sensor emits an ultrasonic pulse, and when it encounters an object, the pulse is reflected to the sensor. By measuring the time it takes for the pulse to return, the sensor calculates the distance to the object

Laser range finders use laser beams to measure distances. The sensor emits a laser pulse that reflects off the target object. By measuring the time it takes for the laser pulse to return, the sensor calculates the distance based on the speed of light.

Infra red (IR) sensors emit infrared light and detect its reflection off a surface or object. By measuring the intensity of the reflected IR light, these sensors can estimate the distance. IR sensors are used in proximity sensing applications, such as object detection in appliances

2.4.2 Object Detection and Localization

Object Detection entails the techniques for identifying and locating objects under observation in a digital image. Object detection techniques function by classifying different objects termed classification and locating the position of the object in the image termed localizations. The classification refers to assigning labels to each detected object while the localization involves drawing a bounding box around each of the detected objects in the image. Most assistive devices for the blind rely on object detection for recognizing the objects in the images captured when using the camera as a visual sensor [32]. We need to review the methods employed in object detection.

Over the years, there are several techniques have been developed for the object detection task with two objectives in mind accuracy and speed []. These techniques are often grouped based on the underlying concepts and design where they are ensembled from hand-engineered approaches called "traditional methods" or "deep learning-based approaches" which employ deep neural networks for the tasks. Several review works in object detection have classified these techniques and we present a summary in Table 2.3.

There are three notable traditional methods introduced in the early development stage of the object detection tasks. These pioneer methods are viola-Jones [37], Histogram of Gradients (HOG) [38] and the Deformable Parts Model (DPM). The Viola-Jones algorithm introduced the concept of Haar cascades for object detection. It provided a robust and efficient method for face detection, making it a breakthrough in computer vision. The HOG algorithm, proposed by Navneet Dalal and Bill Triggs, revolutionized pedestrian detection [38]. It extracted features based on the distribution of gradients in an image and achieved state-of-the-art performance at the time.DPM, introduced by Pedro F. Felzenszwalb et al., extended the concept of part-based models for object detection. It improved the accuracy of object detection by modeling deformable parts and their spatial relationships.

The advent of Convolutional Neural Networks and the development of very deep neural networks for image classification led to a new approach to object detection. The early works to implement CNN for the object detection tasks are R-CNN [39] proposed by Girshick *et. al*, Fast R-CNN [40] and Faster R-CNN [41]. The approaches are categorized as 2 stages object detectors since they perform object recognition in a multistage process. The first stage is using a pre-trained CNN model on a large image dataset as a classifier to extract features as regions that are proposed for classification and also a bounding box regressor is trained to draw bounding boxes around the detected objects. The R-CNN (Region-based Convolutional Neural Networks) approach, proposed by Ross Girshick et al., combined the power of deep learning with object detection. It introduced the idea of using selective search to generate region proposals and then classifying these regions using a CNN, achieving significant improvements in accuracy. Fast R-CNN, also proposed by Girshick et al., improved upon R-CNN by sharing convolutional features across multiple region proposals. This approach eliminated the need for time-consuming region-wise convolutional operations and significantly sped up the object detection process. Faster R-CNN, introduced by Shaoqing Ren et al., further enhanced the R-CNN framework by incorporating a Region Proposal Network (RPN). The RPN generates region proposals directly from the convolutional feature maps, enabling end-to-end training of the entire system.

A further improvement was made to 2 stages to make the detections into one step process for detecting objects. Single-Stage Object Detectors, represented by models like OverFeats [42], YOLO (You Only Look Once) [43] and SSD (Single Shot MultiBox Detector)[44], follow a one-step process for object detection. These types of detectors are faster and more suitable for real-time applications because they eliminate the need for region proposal generation. The one-step process simplifies the architecture, making it easier to train and deploy. They are more memory-efficient, making them suitable for resource-constrained environments.

The field of object detection continues to evolve, with researchers working on hybrid approaches that aim to combine the strengths of both two-stage and single-stage detectors. The choice of which type of detector to use ultimately depends on the application's specific needs and the available computational resources.

We summarize the notable object detection approaches in 2.3 based on the approach or category, the year it was released, and a description.

TABLE 2.3: A summary of the most common state-of-the-art object detection approaches

Category	Name	Reference & Year	Description
Traditional	Viola Jones HOG DPM	[37], 2001 [38], 2005 [45], 2008	Uses hand engineered features combined with machine learning algorithms
Deep Learning Based (CNN based 2 stage)	RCNN Fast RCNN Faster RCNN	[39], 2014 [40], 2015 [41], 2015	Uses regional proposal network for recognizing objects in images
Deep Learning Based (Single stage)	OverFeat YOLO SSD	[42], 2013 [43], 2016 [44], 2016	Uses a single CNN model for object detection

2.5 Common sensory methods used in Assistive devices

To furnish visually impaired persons (VIPs) with information about their surroundings, various sensory devices are employed in assistive technologies. These devices serve to convey essential data regarding object shapes, sizes, distances, and locations in front of VIPs. In the following section, we delve into some commonly utilized sensory devices in assistive technologies.

2.5.1 Infrared-Based Sensors

Infrared-based sensors are a vital component of assistive navigation devices. These sensors emit infrared light and measure the time it takes for the light to bounce back from obstacles, providing crucial distance information. They are particularly effective for detecting nearby objects, making them an essential feature in many navigation aids for the blind. Infrared sensors offer real-time feedback, enhancing safety and obstacle avoidance.

2.5.2 Camera-Based Systems

Camera-based systems are revolutionizing assistive navigation. These devices use built-in cameras to capture live video feeds of the user's surroundings. Advanced image processing converts this visual data into auditory or tactile feedback, providing users with invaluable environmental information. These systems excel in recognizing objects, and text, and even providing navigation guidance, making them versatile tools for the visually impaired.

2.5.3 LiDAR (Light Detection and Ranging)

LiDAR technology is gaining prominence in assistive navigation devices. It uses laser beams to measure distances and create detailed 3D maps of the environment. LiDAR offers high precision, making it effective for detecting both near and distant obstacles. Its ability to provide rich spatial data enhances safety and helps users gain a better understanding of their surroundings.

2.5.4 Radar-Based Systems

Radar-based navigation systems employ radio waves to detect objects and obstacles. While commonly used in aviation and automotive applications, they are finding utility in assistive devices for the blind. Radar can provide real-time information about moving objects and offer an extended range for detecting obstacles. When integrated with auditory or tactile feedback, radar enhances situational awareness and aids in safe navigation.

TABLE 2.4: A summary of the technologies used in developing Electronic Travel Aids with their advantages and disadvantages

Sensor	Transmission mode	Advantages	Disadvantages
Ultrasound	Sound waves above 20 kHz [emitted]	• Low cost, simple to operate • Lightweight, robustness, and fast response time • Good performance in poor lighting and transparent objects • Detect a wide range of materials	• Not suitable for medium to long-distance range, normally more than 5 m • Affected by temperature, pressure, and ambient noise in the environment • Wide beamwidth and sensitivity to mirror-like surfaces cause specular reflections • Cannot distinguish shape and size • Must be perpendicular to the target as possible to receive the correct range data.
Infrared	Infrared Light [emitted]	• High-resolution, low-cost, and lightweight • Faster response time than ultrasound • Can measure temperature	• Sensitive to weather conditions • Short detection range • Affected by dim light conditions
Camera	Visible light [ambient]	• Cheap, Compact size • Rich Contextual information • Vision similar to human eyes • No interference problems with the environment • Estimate boundaries of objects	• Requires ambient light to illuminate the field of view • Susceptible to changes in light, dust, rain, and snow • Requires high Computation cost • No depth information provided
RADAR	Millimeter wave radio waves [emitted]	• Reliable • Accurate measurement for distance and relative speed • For medium to long-distance range (200 m) • 150° wide field of view • Good Angular Resolution • Robust in different weather and environmental conditions	• Expensive • Heterogenous reflectivity of materials makes processing ambigious • Lower processing speed compared to camera and lidar • Lacks fine resolution needed for obstacle detection
LIDAR	Laser Signal [emitted]	• Accurate distance measurement • Wide field of view • Precise measurement of depth • 360° high-resolution mapping • Can measure outlines of objects • Unaffected by lighting conditions	• Expensive • Affected by dust, rain, and snowy conditions • Only objects in the scanning plane are detected • 3D point cloud storage requires large memory • The point cloud is sparse

2.5.5 StereoVision

In this section The technique of using two cameras placed parallel to each other to see the same object is termed Stereovision. The two cameras are separated by a baseline, the distance for which is assumed to be known accurately. The two cameras simultaneously capture two images. The two images are analyzed to note the differences between the images. Essentially, one needs to accurately identify the same pixel in both images, known as the problem of correspondence between the two cameras. Features like corners can be easily found in one image, and the same can be searched in the other image. Alternatively, the disparity between the images can be found to get the indicative regions in the other image, corresponding to the same regions in the first image, for which a small search can be used. The disparity helps to get the depth of the point which enables projecting it in a 3D world used for navigation.

For the application, the cameras should have the same focal length as their X-axis intersecting and aligning with the baseline. Fig. 3.8 illustrates a typical stereo-vision system

Application of object detection in mobility aids for the visually impaired is quite common and simple to user [46].

2.6 Related Works using the various technologies

Numerous devices have been developed to assist visually impaired individuals in various mobility tasks, with a particular focus on mobility. In this section, we will explore and categorize these devices, providing comprehensive insights into their functionality and technology.

These categorizations are based on technological similarities and enhancements introduced by various studies. These devices often employ different sensing components, with imaging sensors like cameras being the most prevalent. The commonly utilized technologies encompass white canes, imaging sensors, distance measurement sensors, and location and positioning sensors such as GPS. It's worth noting that some Electronic Travel Aids incorporate multiple sensing components for enhanced functionality.

TABLE 2.5: List of categorizations of various ETA reviewed

Category	Description of Category
Canes Based	The use of a walking stick painted white
Echolocation Based	The use of sensing device for detecting obstacles
Belt Based	Attachment of sensing device to a belt worn by the VIP
Camera Based	The use of image sensing devices such as camera for detecting obstacles

2.6.1 White canes and their improved modifications

The white cane is one of the oldest means of guiding the VIP with mobility tasks. It was first introduced by James Biggs of Bristol in 1921. He claimed to have invented the white cane after an accident claimed his sight, the artist had to acclimatize to his environment. Feeling threatened by increased vehicular traffic around his home, Biggs thought of painting his walking stick white to make himself and his condition more conspicuous to motorists. Ten years later, the white cane became established its presence in society [47].

There are several identified challenges with the whitecane and studies have been undertaken to improve it by adding other sensors and peripheral devices to enhance the features and make it more convenient for them

Laser Cane [48], Teletact [49], and Minitact [50] are laser augmented canes. These canes are useful for detecting floor-level to head-level obstacles in front of them. A subject needs to continuously scan the surrounding environment as these canes use very narrow beam laser devices. The distance measurement in these devices is susceptible to interference due to natural light. The laser cannot detect transparent glass as its beam traverses through the glass without being reflected. Furthermore, the high cost and the significant expertise required to operate these devices are a major concern. Tom Pouce [8] and RecognizeCane [10] refer to the augmentation of a cane with infrared sensors. The applicability of these canes in the outdoor environment is limited due to their infrared sensors.

K-Sonar Cane [11] and Ultracane [51] [52] are whitecanes with ultrasonic sensor-augmented devices. K-Sonar Cane uses low-pitch and high-pitch sounds to convey the distance of obstacles. It requires good scanning and sound interpretation skills. Ultracane assists in detecting floor-level planes to obstacles at almost the height of the users in the traversable path. This group of canes requires complex techniques for their usage same as the traditional white cane.

GuideCane [53] is also an ultrasonic sensor-augmented but designed as a robotic guiding cane that is rolled on passive wheels to support the weights during walking. The GuideCane can detect floor planes located at front-way and sideways obstacles. This cane is large in size and has a limited scanning area [6]

2.6.2 Echolocation Based

Echolocation, a technique initially employed for the benefit of visually impaired individuals, was pioneered in the development of mobility aids [54]. These echolocation devices utilize ultrasound signals with frequencies ranging from 70 to 40KHz, similar to the ultrasound range used by bats for their echolocation.

In their early experiments, these devices aimed to assess a user's ability to discern obstacles in their immediate vicinity. Several notable devices emerged, including the CyARM [55], miniguide [56], sonic torch [57], and kaspa system [58]. These handheld echolocation-based devices were designed to detect obstacles in front of the user, but they necessitated continuous manual movement in multiple directions for consistent obstacle detection.

The CyARM device introduced the use of an ultrasonic transducer to help visually impaired individuals spatially localize themselves within their surroundings. This device measures the distance between the user and obstacles using an ultrasonic sensor and conveys this distance information to the user through a haptic feedback sensor. In real-world outdoor scenarios, the system demonstrated high detection rates for stationary obstacles, accurately estimating the distances between users and obstacles. However, it encountered challenges when detecting dynamic obstacles, experiencing a noticeable decrease in accuracy and recall rates, by up to 30%. Additionally, the performance of ultrasonic sensors could be adversely affected by various weather conditions.

2.6.3 Belts

These are designed to be worn around the waist with several sensory devices attached to them to provide feedback to the users.

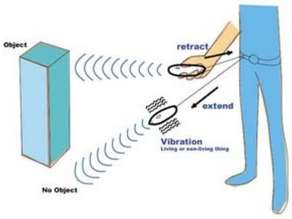


FIGURE 2.1: The proposed prototype of the CyARM

Navbelt [59], ultrasonic waist-belt [60], and electronic bracelet [61] are wearable assistive aids that are implemented in the shape of belts and bracelet. These aids are limited in detecting floor-level obstacles. Echolocation system for the blind (ESB) uses a phase beamforming approach to perceive the surrounding environment [62]. It is a promising system, but its complexity can be a major concern. Binaural sonar ETA [63] is a small wearable device that can be used for landmark recognition, obstacle motion perception, and texture recognition. The field evaluation details of this prototype aid are not provided. A navigation system for the blind [18] is a multisensory system for augmenting blind navigation. The information content of this system is less and it works under the condition of low noise surroundings only.

[64]

2.6.4 Camera Based

In the realm of assistive technology for the visually impaired, camera-based systems have played a significant role in transforming visual information into accessible formats. Below, we explore several notable camera-based assistive devices designed to enhance the independence and safety of visually impaired individuals.

The vOICe, [65] is a video-guided assistive system for converting visual information into different sound patterns for the visually impaired. The vOICe systems comprised of a camera to capture images, and a computer that uses the system to map the learning curve for these sound patterns is quite steep [6].

The BrainPort vision device uses a tongue display unit to assist the subject in navigating around obstacles proposed by Arnoldussen [66]. It translates visual information acquired from a camera into an electrical stimulation pattern and displays this pattern on the tongue of the subject via a 20 x 20 electrode array. A clinical evaluation study of this promising device has been scheduled.

A stereo vision aid [67] is a multisensory system that employs disparity measurement for obstacle detection and uses motion estimation and inclinometer for overall understanding of the surrounding environment. This system is expensive and large.

A navigation assistant designed by Tapu *et. al* [68] a navigation assistant, consists of a smartphone attached to a chest-mounted harness. It employs a multifaceted approach to detect obstacles, including points of interest tracking, motion estimation, background motion estimation, and object classification. This system aids partially sighted individuals in autonomous navigation and has plans for improving its alerting capabilities.

Tapu *et. al* [69] DEEP-SEE framework integrates computer vision techniques and deep convolutional neural networks (CNNs) to detect, track, and recognize objects in real-time during outdoor navigation. It encompasses object detection, motion-based object tracking using CNNs, and predictive location modeling based on visual similarity. The device aims to enhance cognition and safety for visually impaired individuals navigating complex urban environments.

2.6.5 Limitations and Challenges with various ETA devices

The limitations of existing ETAs and the mobility difficulties of visually impaired subjects motivated us to research assistive technology. The objective of this research is to develop an ETA that can (1) construct a logical map of the surrounding environment, (2) interpret, categorize, and prioritize situation-specific details of the environment to reduce information overload, and (3) simplify the representation of the priority information. Addressing limitations of existing ETAs and requirements of visually impaired subjects we developed a novel aid called electronic mobility cane (EMC) at the Indian Institute of Technology (IIT) Kharagpur, India. It was named VENUCANE [70]. The EMC has remarkable distinctions to serve the mobility requirements of visually impaired people. It constructs the logical map of the surrounding environment and interprets the distribution of obstacles by relevance of their distances from each other. It simultaneously detects multiple obstructions namely sideway obstacles, front-way obstacles, and the information about the floor status without the user's additional perceptual effort. It assists in proactively detecting ascending staircases, floor-level obstacles, knee-level obstacles, waist-level obstacles, trunk-level obstacles, head-level obstacles, blocked front-way, left turns, right turns, and blocked all-sides situations. It categorizes detected obstacle situations and deduces the priority information from them. It represents the distance of the obstacle in terms number of walking steps of the subject and maintains a safety margin distance to negotiate nearby obstacles. It conveys a simplified representation of the perceived information to the subject by using intuitive vibration, audio, or voice feedback. The EMC system is available in both wired and wireless modes of operation. The wireless configuration is proposed to further minimize the physical interface of the EMC with the natural sensory channels of the subject. This paper covers the design, development, and clinical evaluation of the EMC.

2.7 Conclusion

In this chapter, we discussed literature on various areas that are related to our research. These include vision impairment, the causes of vision impairment, and the impact that it has on individuals who lost their sight. A further review is conducted on various devices that have been proposed and developed as aids for the visually impaired. These devices has several benefit and also come with some limitations based on the technologies and sensory devices being used.

Chapter 3

Wearable Assistive Device for Navigation

3.1 Introduction

In this chapter, we provide an in-depth description of our specialized wearable assistive navigation devices called WAND, designed with a focus on assisting VIPs with navigation and mobility. The overall details of architectures, algorithmic frameworks, and fundamental blueprints that form the foundation of these devices are presented in this section. Our wearable assistive device is built upon four fundamental modules: obstacle avoidance, ground plane detection, pathfinding, and speech communication.

We grouped the four modules into two categories based on the types of input modal the module works with. The visual category is for the modules that work with images while the Audio category is for the modules that work with the audio modal. The categorization is illustrated in Figure 3.1

We delved into the visual aspect of our system, encompassing obstacle avoidance, plane detection, and pathfinding. In parallel, we also explored the audio component, a speech communication system that included functions such as speech recognition, speech enhancement, and speech synthesis. Throughout that chapter, we elucidated the functions of our system and underscored its innovative aspects, particularly its capability to seamlessly execute these functions simultaneously. The visual system primarily focused on tasks like object recognition and scene detection, especially in the user's frontal field of view. Its core objective was to promptly alert users to potential obstacles along their path. Meanwhile, the audio system's primary role was to notify the user and handle any requests they might make.

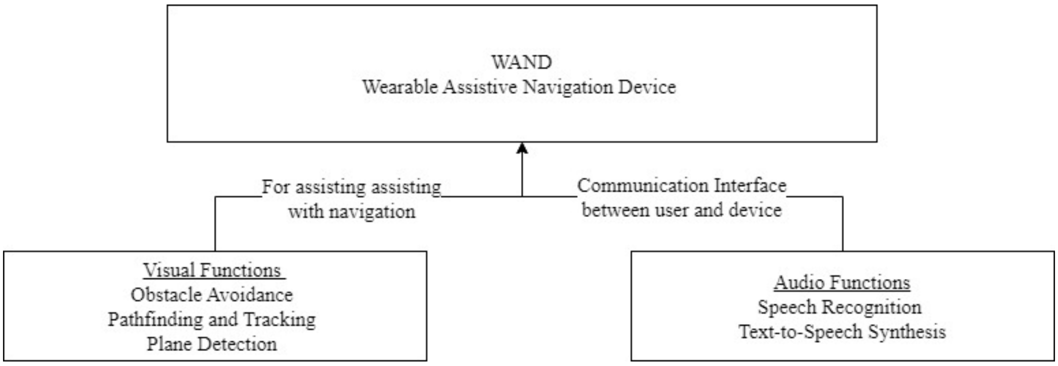


FIGURE 3.1: An illustration of the categorization based on the input modalities of the modules integrated into the assistive device.

3.2 The wearable assistive navigation device, WAND

WAND is the term we coined for our proposed wearable assistive navigation device, which not only aided in navigation but also facilitated simple requests such as asking for guidance to a destination by the VIPs. A feasibility study was conducted to gain insights into the specific needs of visually impaired individuals, particularly regarding their awareness of the immediate environment, orientation, and obstacle avoidance. This study revealed that having a device that provided an audio interface for communication and feedback would appeal more to the VIPs.

Our literature review of existing approaches affirmed that this research was both feasible and socially valuable. It held the potential to empower the visually impaired by granting them a degree of freedom and independence. However, achieving this overarching goal was a complex endeavor, necessitating a breakdown of the primary objective into smaller, manageable sub-goals that could be systematically addressed and eventually integrated to create an efficient and effective device. Therefore, we had four modules for the sub-goals: the obstacle avoidance module, the pathfinding and tracking module, the ground plane detection module, and the speech communication module. Below, we outlined the key objectives that guided the development of this device:

- To assist the VIP in moving from one location to another by safely navigating to their destination.
- To formulate a robust and dynamic obstacle avoidance strategy for the blind.
- To check the ground plane of the user to ensure they are walking on safe ground.
- To respond to the simple requests that VIPs make about their destination and environment.

The proposed system was designed to have modules that could effectively handle the objectives while ensuring there was a synergy between them to perform efficiently in assisting the user to navigate through their environment, whether it was indoor or outdoor. The flow chart presented in Figure 3.2 shows how the system operating the device would process the data disseminate the information and alert the user on the various functionalities that were in the device.

3.2.1 Navigation Solution in WAND

A navigation system aims to provide users with helpful information to the user to get to a destination point from a given point safely while monitoring the user's position in a modeled map. Researchers have been working in this field to find highly optimized safe, and cost-effective solutions to guide the VIP in both outdoor and indoor environments. The navigation focuses on the location of the VIP and getting them to their desired destinations. To achieve this, we considered three functionalities we combined to ensure the VIP is able to reach their destination safely. This includes avoiding obstacles, avoiding and taking note of discontinuities in ground planes and pathfinding with tracking .

3.2.1.1 Obstacle Avoidance Solution

Obstacle avoidance is one of the most important modules or components in an ETA system. The ability of the VIP or ETA users to do a self-evaluation of how to avoid an

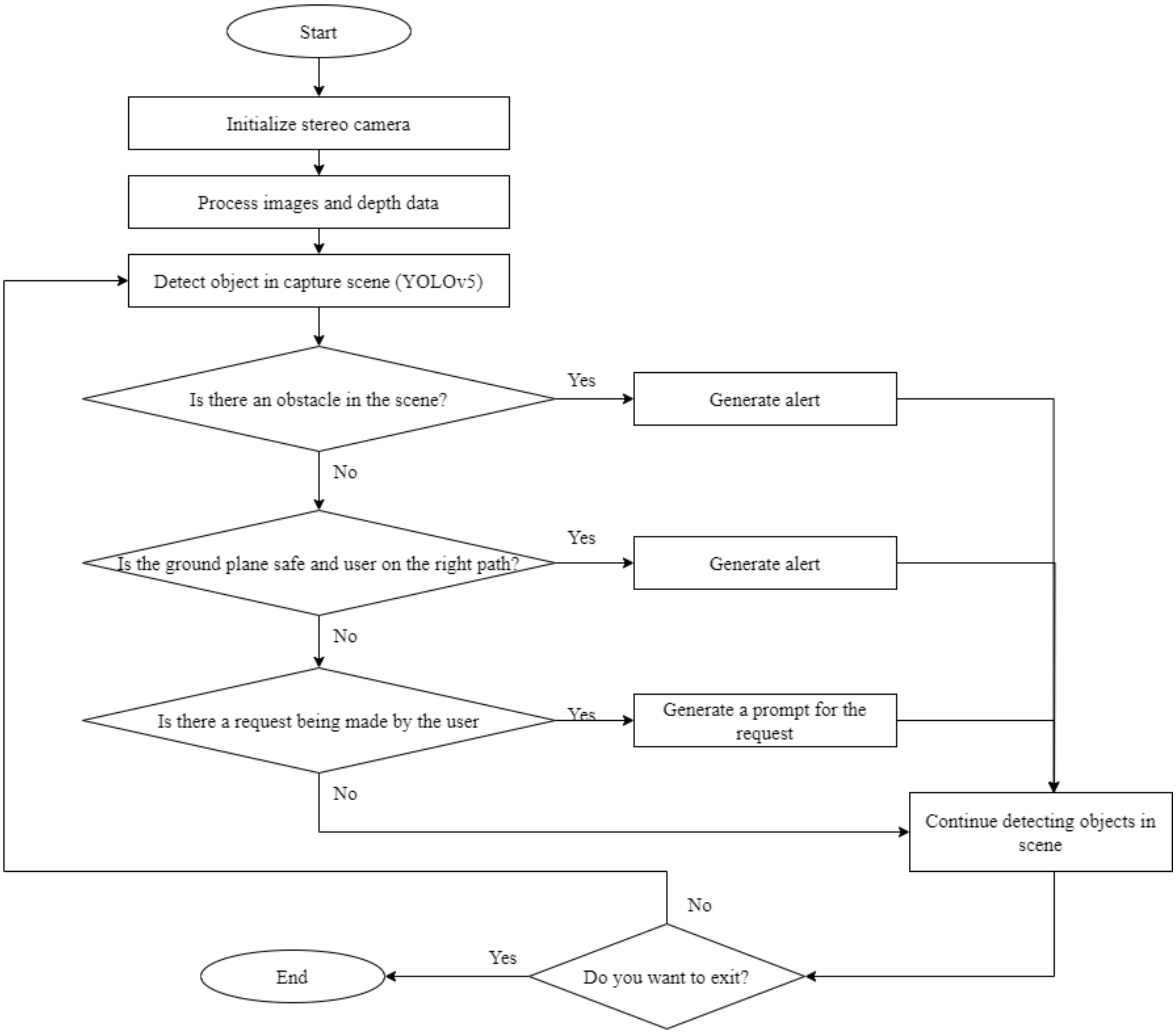

Start
Initialize stereo camera
Process images and depth data
Detect object in capture scene (YOLOv5)
Is there an obstacle in the scene?
Yes
Generate alert
No
Is the ground plane safe and user on the right path?
Yes
Generate alert
No
Is there a request being made by the user
Yes
Generate a prompt for the request
No
Continue detecting objects in scene
No
Yes
End
Do you want to exit?

FIGURE 3.2: A flow chart of the various goals integrated with the flow of information, and alerts with the device

obstacle in their proximity depends on the information provided by obstacle avoidance module systems. Obstacle avoidance is the first safety feature that assistive devices for the VIP such as ETA is required to have to be functional. Obstacle avoidance is to ensure the safe and efficient navigation of a system or individual through an environment by identifying potential hazards and determining appropriate strategies to steer or move away from obstacles.

We presented our approach using a process flow in Figure 3.3. The process flow show that the processes involved in identifying obstacles in front of the user. Information about the obstacles is updated in real time after each frame is streamed from the camera. The system stores information about possible obstacles and the changes that occur to their state in terms of being stationary or in motion.

In addressing mobility and navigation tasks, motion plays a pivotal role, and a substantial focus is placed on understanding the implications of various types of motion within these tasks. For instance, when it comes to obstacle avoidance, it's essential to consider not only the motion of the user but also the motion of obstacles, whether they are approaching or receding from the user. In certain scenarios, the speed of these movements becomes a crucial factor in obstacle avoidance strategies. The swiftness or sluggishness of these motions can influence the user's reaction

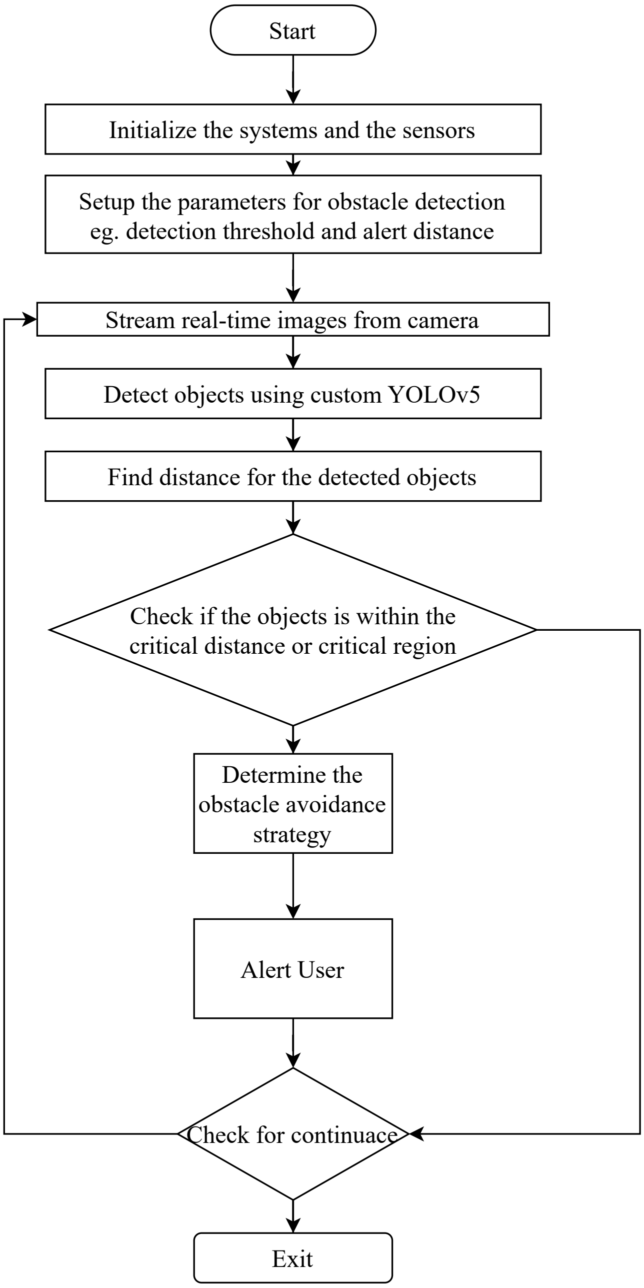

Start
Initialize the systems and the sensors
Setup the parameters for obstacle detection eg. detection threshold and alert distance
Stream real-time images from camera
Detect objects using custom YOLOv5
Find distance for the detected objects
Check if the objects is within the critical distance or critical region
Determine the obstacle avoidance strategy
Alert User
Check for continuace
Exit

FIGURE 3.3: A flowchart illustration of the obstacle avoidance process implemented in this wearable navigation system.

time, particularly in terms of coming to a stop. In the realm of physics, this reaction time required to stop is known as "braking distance," a fundamental concept with widespread applications across various fields.

Braking distance is defined as the distance an object in motion travels before coming to a complete stop. This distance is contingent on numerous factors, with one of the most critical being the object's kinetic energy. The concept of braking distance finds applications in diverse areas, including transportation systems [71], vehicle performance metrics, and the design of braking systems.

In the context of walking, humans possess kinetic energy, and as a result, they require a certain amount of time and distance to come to a full stop. This stoppage might be prompted by the detection of an obstacle in their path or a decision to make a turn. However, for visually impaired individuals, the ability to perceive when to stop or change direction due to an obstacle is absent. Hence, in the development of an obstacle avoidance system, the concept of "time to contact" becomes crucial.

We formulate the time-to-contact based on the equation 3.1.

$$\tau = \frac{Z}{v} \tag{3.1}$$

Where Z is the distance between the camera obstacles, and v is the velocity of the camera with respect to the obstacles. We can assume that the velocity of the camera is the same as the walking pace of the user. We determined v using the difference in distance of the obstacles from the camera at t_1 and t_2 since the user is consider to be moving and the object or obstacle may also be in motion so the most accurate change per the duration with the change in the distance between the obstacle and the camera.

$$v = \frac{\triangle Z}{t_2 - t_1} \tag{3.2}$$

Even though time to contact is known to be a very effective means of determining how to alert the VIPs of the obstacles in their path, in this approach we go further to let them know how close the objects are to them and within two (2) meters. We consider objects within the 2-meter parameter to the VIP as very hazardous and the possibility of bumping into them. For this reason, we have the assumption that the time of contact must be at least equal to a reaction period of two (2) meters. A detailed description of the innovation of the obstacle avoidance strategy are detailed in Chapter 4 of the thesis.

3.2.1.2 PathFinding and Tracking

The WAND system offers an indoor and outdoor approach to navigation using both computer vision and a GPS based techniques. It integrates stereo camera data for visual odometory for indoor environment and a GPS sensors for outdoor environment to track user position and provides feedback via speech alerts if the user moves away from the tracks. Figure 3.4 shows an sequence of action or activities that are in pathfinding and tracking module of the WAND devices. This to provide an effective and safe way for the VIP to navigate to their desired destinations.

The interaction between a VIP , a GPS system, a pathfinding algorithm, and the destination in the context of navigation as illustrated in Figure 5.5. The user's current position is determine with a GPS sensor in an outdoor environment, also their desired destination is capture from the command request they make and the information is forwarded to the GPS system. The GPS system then sends these coordinates to the pathfinding algorithm, which calculates the most efficient path to the destination. This optimal path is sent back to the GPS system, which prepare the outdoor for the speech communication module to inform and update the user about their path. As the user moves, the GPS system continuously tracks the movement in a loop, updating the user's location in real-time and adjusting the path as necessary to account for any deviations or changes in the route. This loop continues until the user reaches their destination, at which point the process concludes. The proposed method depicted by sequence diagram illustrates the dynamic nature of using GPS built into navigation systems, which can adapt to changes in the user's location and provide real-time updates to ensure they reach their destination efficiently.

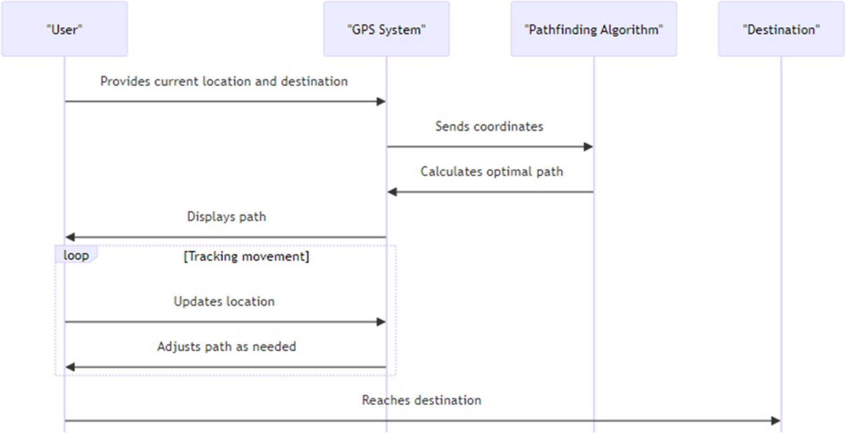


FIGURE 3.4: An illustration of a sequence diagram depicting timelines various phases in supporting the user to navigate from their current position to their destination

3.2.1.3 Ground Plane Detection

Plane detection is very important since it helps identify ground changes in front of the visually impaired to inform them of unleveled ground. Ground changes give the VIPs a lot of challenges since they cannot perceive the changes and how to overcome them. It is known that the VIP are not able to adjust to the changes in the ground plan using a natural mechanism called sway [72]. This incapability of adjustment of posture often leads to higher risks of falling.

It is therefore prudent to consider developing a plane-checking module within the obstacle avoidance system to inform the user of the changes in their ground plane. The module should check if the planes are horizontal enough for the user to normally continue walking or if the user needs to adjust their posture to the changing ground plane.

3.2.1.4 How the Ground Plane Detection Module Works

The ground plane detection is based on using a 3D point clouds generated from the depth data. A 3D mesh of the points that satisfy the plane equation 3.3 are constructed and compared with a normal vector of the world cordinate $\vec{n}$ which is the y-direction to be the vertical to determine the orientation of the plane. The normal can be calculated by find the cross product of two vectors on the plane. For instance, given vectors $\vec{p}$ and $\vec{q}$ on the same plane, the normal vector will be $\vec{n} = \vec{p} \cdot \vec{q}$.

$$ax + by + cz = d \tag{3.3}$$

Figure 3.5 illustrates the flowchart of the ground plane detection flowchart. The process starts when the stereo camera is initialized to set up the configurations of the camera. A reference point within the captured mesh is initialized and then a plane equation is set to select points with the mesh that satisfy the plane equation which is presented as equation 3.3. The normal is used to determine if the obtained plane is vertical or horizontal. If the plane is not horizontal, the VIP or the user is informed through the system alert about the irregularity in the plane.

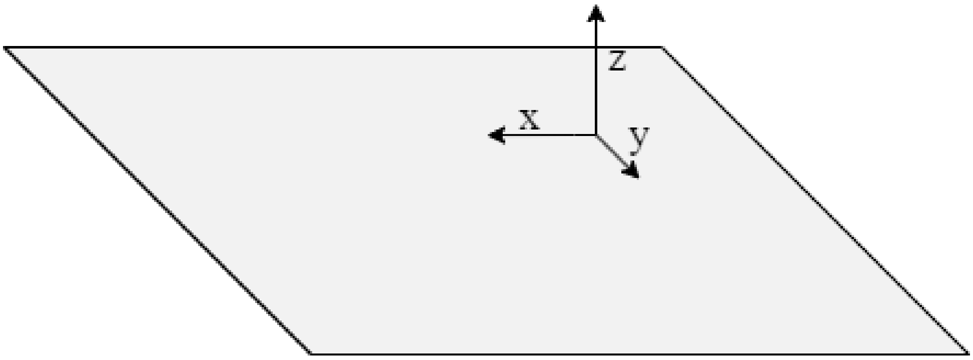


FIGURE 3.5: An illustration of the plane a under consideration the direction of the expected normal

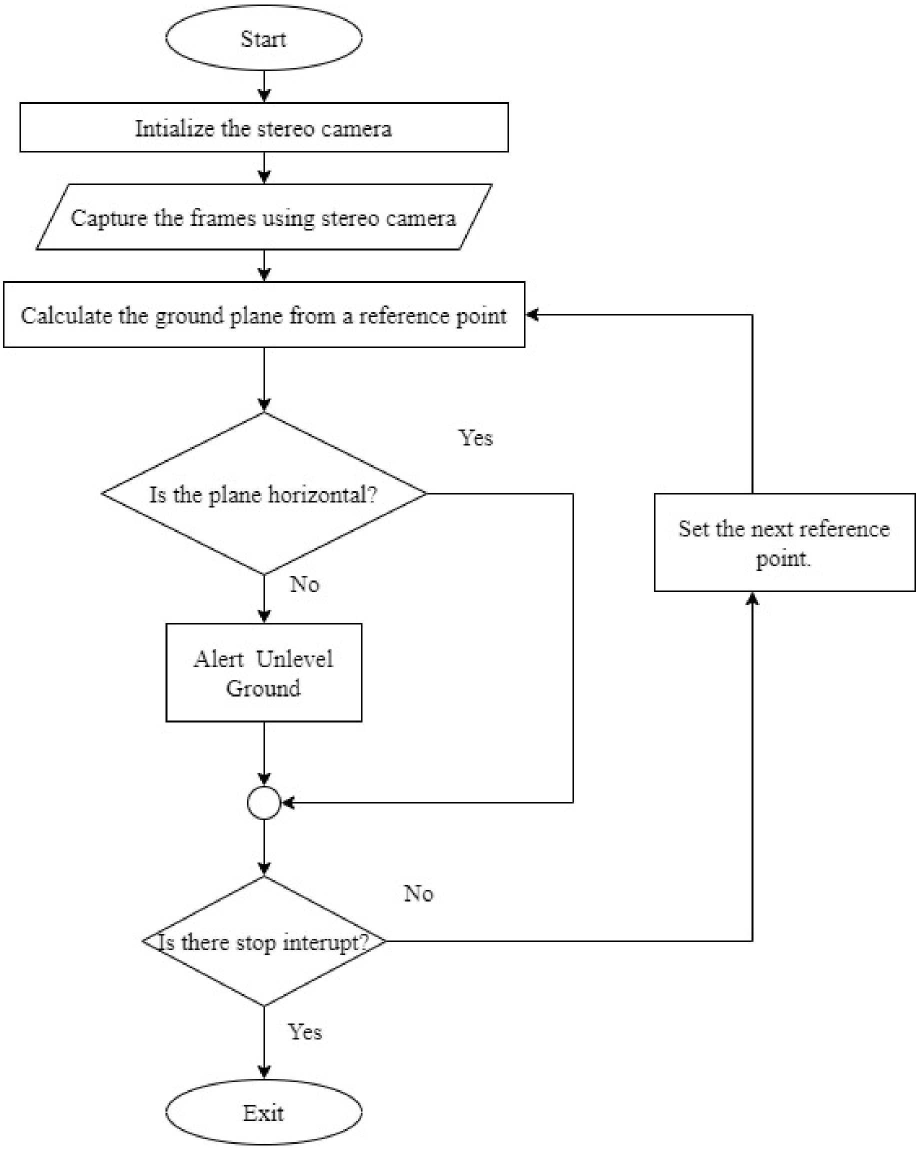

Start
Intialize the stereo camera
Capture the frames using stereo camera
Calculate the ground plane from a reference point
Is the plane horizontal?
Yes
No
Alert Unlevel Ground
Set the next reference point.
Is there stop interupt?
No
Yes
Exit

FIGURE 3.6: A flowchart illustration of the ground plane

3.2.2 Audio Communication

Designing effective feedback systems for wearable assistive devices is crucial as they serve as the interface between the device and the user, particularly for ETAs for visually impaired persons VIPs. Common feedback mechanisms include auditory and tactile approaches. In this context, we propose a novel approach rooted in the neural perception concept of dorsal and ventral pathways.

There are several schools of thought on visual perceptions have considered two visual streams (the dorsal stream/pathway and the ventral stream/pathway) [73] [74] [75]. The ventral system focuses on 'what' whereas the dorsal system focuses on 'where'. The ventral system is connected with recognition and the form representation of objects captured in a scene by the eye. The dorsal stream is associated with the spatial and motion details of the objects in a scene. This thought helps us to choose visual information that is beneficial to us in terms of the tasks humans or animals are undertaking. For instance, mobility tasks would require having information if the path is clear and there are no blockades. The task of grasping would require extra information about the size, texture, or motion of the object. This function thus makes the vision system an important feature of the human body.

The biologically inspired systems can be found in our daily lives. One such inspired system is the computer vision system. The eye and the brain performing the visual perceptions have been modeled in various computer vision systems for robotics, self-driving cars, drones, and mobility aids for the blind as well. The more we understand the relationship between the eye and the brain performing the visual perception tasks the better this system also becomes.

The dorsal pathway, responsible for spatial processing and action guidance, can enhance navigation systems by providing real-time user position updates, direction cues, and obstacle alerts. On the other hand, the ventral pathway, associated with object recognition, aids in identifying landmarks and route confirmation. An effective system integrates both dorsal and ventral processing, combining spatial data with visual recognition cues. For instance, it can audibly instruct users to take a left turn in a specified distance while simultaneously visually displaying a distinctive landmark as a reference point. This integration also enables dynamic route adjustments based on real-time environmental data and user movement patterns, optimizing travel efficiency.

A detailed presentation of how the audio communication is presented in Chapter 5. In that section we talked about the audio module concept and functionalities in providing feedback from the device to the user as well receiving request through speech input from the user and processing to give a response for the input.

3.3 Design and Implementation of the wearable assistive navigation system.

In designing the wearable system, we aimed at a device that could provide the functionalities of ETA with navigation capabilities and thus can perform pathfinding tasks, and obstacle avoidance tasks with command requests such as responding to questions requested by the VIP can be responded to by the device. The system should simultaneously perform all these functionalities since it will give the user a sense of safety and reliability to have assistance.

3.3.1 Hardware Components

The hardware component of our wearable assistive device comprises a single-board computer, Jetson Nano which serves as the central processing unit of the system, and a stereo camera system that provides depth perception, Zed2 Camera [76]. The Jetson Nano is a powerful single-board computer that can serve as the brain of the system. It offers an integrated 128-core Maxwell GPU, quad-core ARM A57 64-bit CPU, and 4GB LPDDR4 memory, along with support for MIPI CSI-2 and PCIe Gen2 high-speed I/O. The ZED2 camera is a stereo camera system that provides depth perception. It can capture RGB images and generate a depth map of the environment. The Zed2 camera is mounted on the microcontroller system to capture the real-time visual input of the surrounding environment. The Zed2 stereo camera uses the disparity calculated from the left and right images to determine the depth information obtained from the camera. The connection of the various components is presented in Figure 3.7.

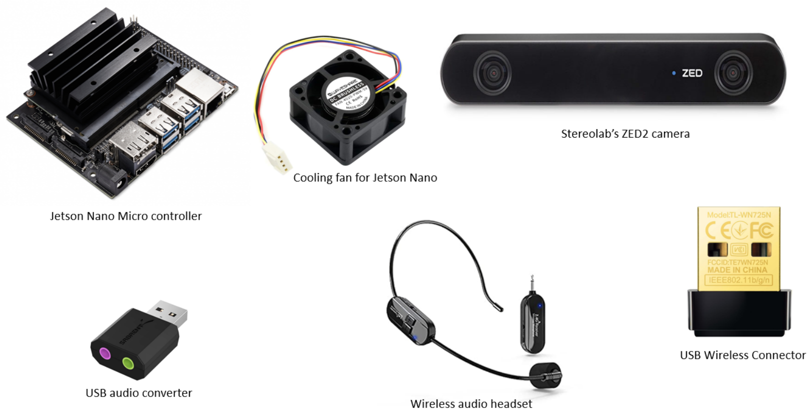


FIGURE 3.7: Various electronic components that were connected to design the wearable assistive device.

The ZED2 stereo camera captures the front view scene of the user. The camera is strategically strapped on the chest to have the front-view scene capture the floor and focal area of 4 meters squared at a distance of 2 meters. The strap with the camera is presented in Figure 3.8 The camera is connected to the Jetson Nano microcontroller. The Jetson is considered a miniaturized computer with the capability of running AI models for the purpose of image classification, object detection, and speech processing. The operating system running on the Jetson is the Linux Jetpack, a special SDK from NVIDIA for supporting the development of end-to-end accelerated AI applications. The wearable assistive system running on the Jetson is written in Python with different modules running simultaneously to guide the user to their desired desired destination.

3.3.2 Software Component

Our system runs on NVIDIA JETPACK 4.1 LTS SDK which is the most comprehensive solution for building AI applications. It includes the latest Linux Ubuntu OS images for the Jetson Nano microcontroller

The obstacle avoidance module comprises a custom YOLOv5 which is integrated with the ZED2 camera SDK module that is responsible for detecting objects in the scene and our developed algorithm then uses the data on the detected objects for

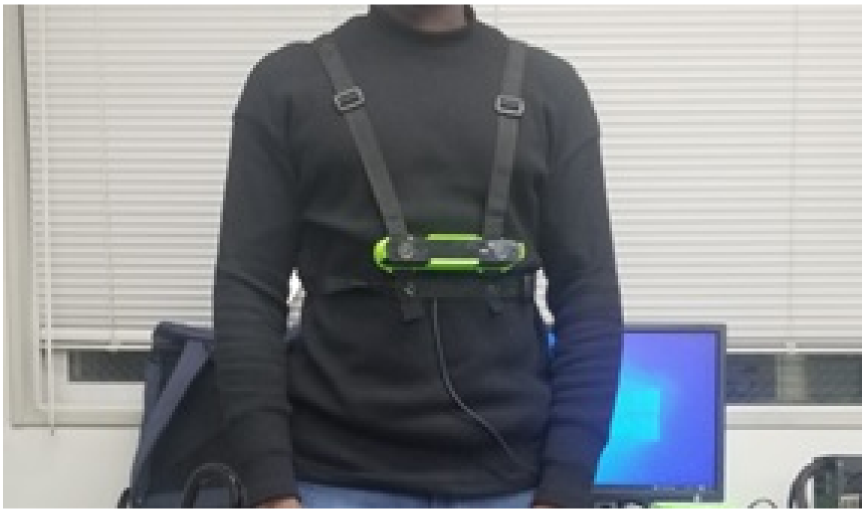


FIGURE 3.8: An image depicting the ZED2 camera being strapped around the chest to capture the front view scene

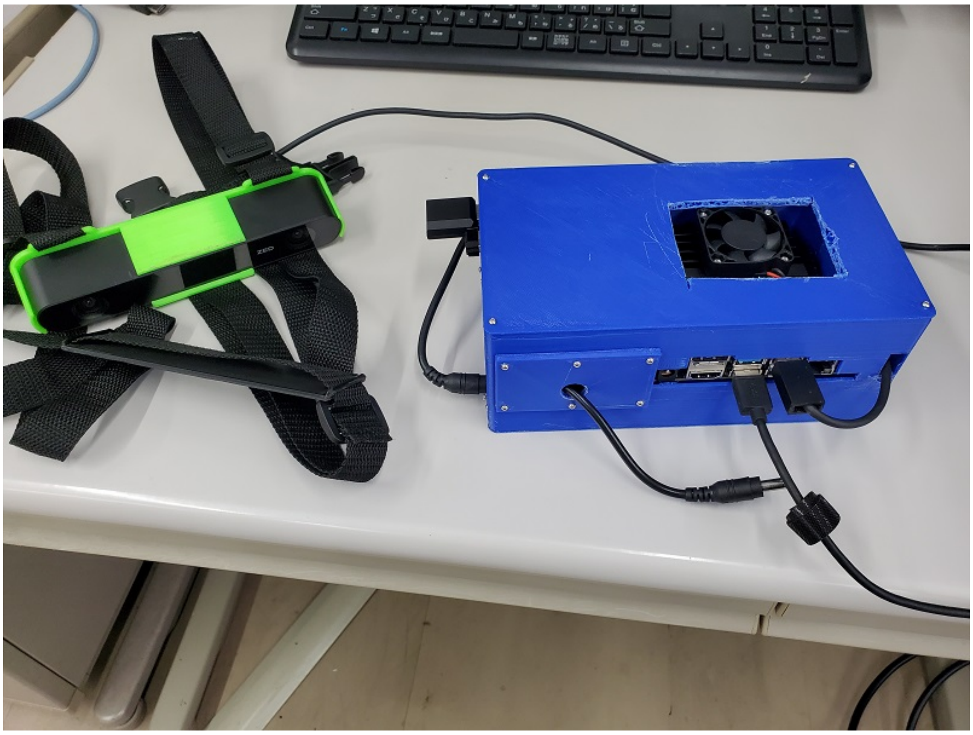


FIGURE 3.9: An image of the 3D printed case holder for the Jetson Nano holder

obstacle selection and avoidance. The obstacle selection and avoidance algorithm is based on object detection therefore a real-time and accurate object detector algorithm is needed to ensure the obstacle avoidance module functions well.

The ground plane detection module uses the 3D mesh data to calculate the planes on which the user walks.

The navigation module is for choosing the optimal path and tracking the user to see if they are on the right path to their destination. This module relies on the Inertial Momentum Unit sensors and other sensors with the ZED2 camera to support locating the position to track the user position using Visual Positioning.

All of the modules run simultaneously and are coordinated such that the alerts for hazardous situations are reported in a single pattern for VIP users to follow as a guide.

3.3.3 Results from the Implemented System

We presented how the system works and how the detection of objects and planes are produced in this section. We show with images the detection of the obstacles and the ground plane detection.

Figure 3.10 and Figure 3.10 depicts a moment before the detection of objects to the moment of selecting an obstacle. Various objects are recognized and then the adaptable grid algorithm helps in determining the objects that are obstacles, then the obstacle that is closest directly in the path of the user is chosen. In this scenario, the table on the right is the imminent obstacle that the user needs to avoid even though the chairs, and tables are detected as objects. Figure 3.11 and Figure 3.12 depict an RGB image and its corresponding depth representation, both derived from stereo images, with the algorithm in action to identify obstacles and determine safe areas for the user's passage. In the color image, you can observe the adaptable grid, which plays a crucial role in obstacle selection, contributing to a faster process. On the right side of the color image, you can see an aerial representation of the detected objects and the tracking of potential obstacles. The depth representation employs 3D bounding boxes to calculate the distance of the detected objects. In the screenshot, the adaptable grid is employed to identify safe regions where there are no overlaps with the detected obstacles. These unoccupied grid cells indicate safe areas, and the user is subsequently directed to shift either left or right to navigate through these safe zones, thereby avoiding collisions with obstacles.

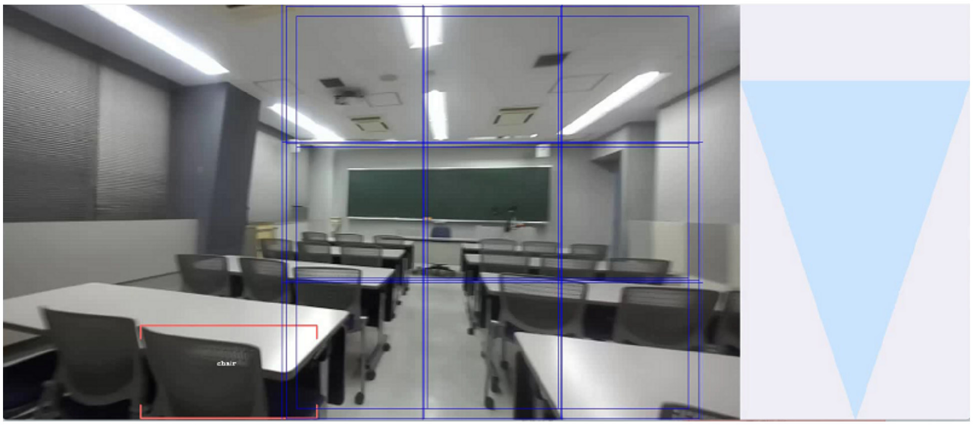


FIGURE 3.10: A screenshot showing the multiple adaptable grid boxes indicated with blue lines created for each detected object and used in determining the object in the proximity of the user to be considered obstacles.

The ground plane checking module checks to ensure the plane is horizontal from the 3D mesh generated from the 3D point cloud obtained from the captured. The detected planes are colored yellow or red if the plane is horizontal or vertical respectively. The plane is colored blue if the system cannot determine if it is either vertical or horizontal. In Figure 3.13 and Figure 3.14 we show how the system updates the plane search and markings.

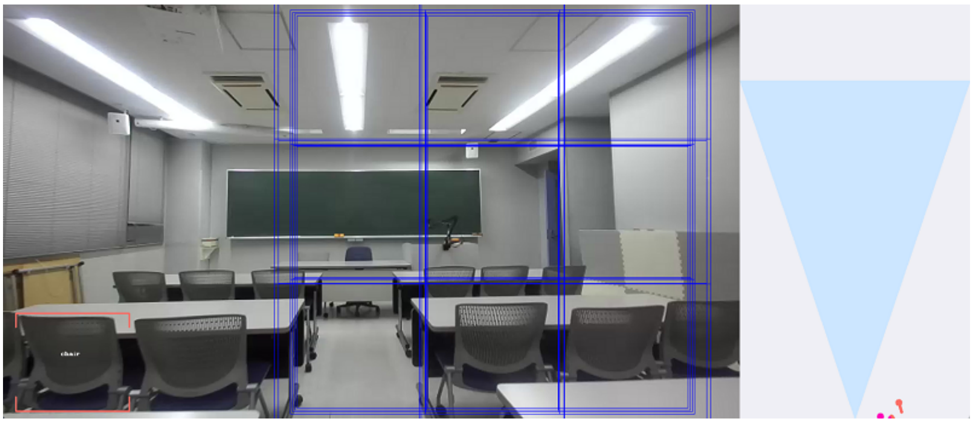


FIGURE 3.11: A screenshot showing the multiple adaptable grid boxes created for each detected object and used in determining the object in the proximity of the user to be considered obstacles

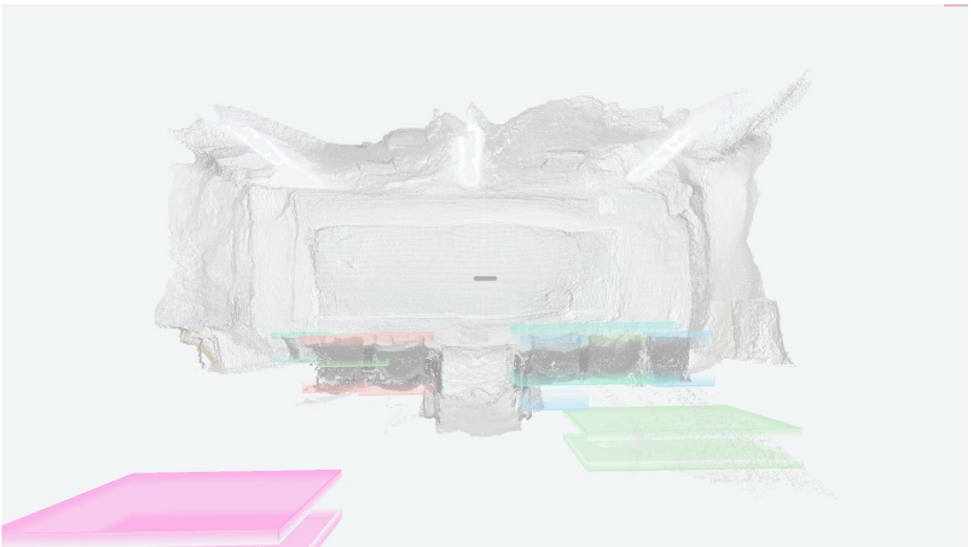
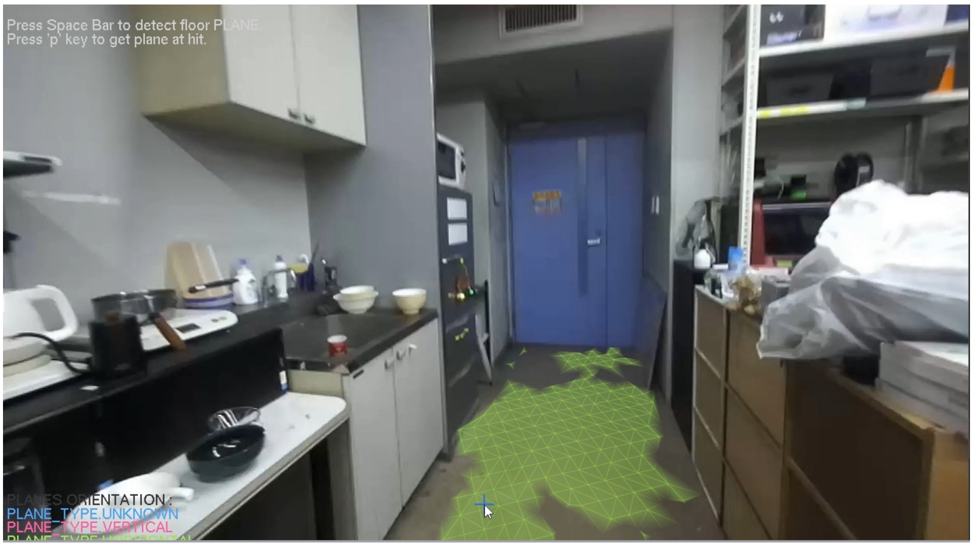


FIGURE 3.12: A screenshot showing the multiple adaptable grid boxes created for each detected object and used in determining the object in the proximity of the user to be considered obstacles.

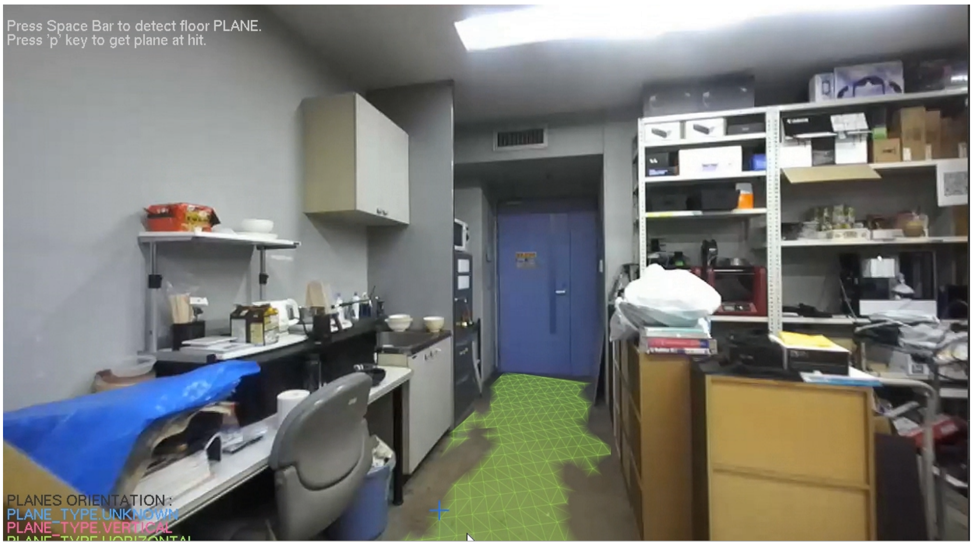
In Figure 3.13 after generating the plane, the next plane is calculated from the new reference point obtained from the current plane to produce the plane shown in 3.14.



Press Space Bar to detect floor PLANE.
Press 'p' key to get plane at hit.

PLANES ORIENTATION:
PLANE_TYPE.UNKNOWN
PLANE_TYPE.VERTICAL
PLANE_TYPE.HORIZONTAL

FIGURE 3.13: The ground plane extracted and overlay on the RGB image at step 1. The yellow mesh shows the safe area for the VIP to walk on



Press Space Bar to detect floor PLANE.
Press 'p' key to get plane at hit.

PLANES ORIENTATION:
PLANE_TYPE.UNKNOWN
PLANE_TYPE.VERTICAL
PLANE_TYPE.HORIZONTAL

FIGURE 3.14: The ground plane extracted and overlay on the RGB image at step 2. The yellow mesh shows the safe area for the VIP to walk on

3.4 Conclusion

We discussed various designs and implementations of our proposed solutions. The design comprises different modules that are integrated into one system to provide various functionalities that meet the navigation needs of the VIP. The wearable assistive device's camera is strapped on the chest of the user to cover the front view path of the user to capture objects in front of the user, and detect the object using a custom YOLOv5 running on a Jetson Nano. An obstacle avoidance strategy discussed in Chapter 4 is used in selecting the potential obstacles among the objects. The user is informed about the obstacle in front of them.

Chapter 4

Robust Obstacle Avoidance and Navigation

4.1 Introduction

In this chapter, we present our novel obstacle avoidance and navigation module in detail including the algorithm formulations, experimentations, results, and discussions. The proposed obstacle avoidance strategy checks the obstacles within close proximity of the user and determines which action needs to be taken to avoid colliding with the obstacle. The experimentations we performed to evaluate the performance of the innovation. The result that we obtained was discussed and observations were highlighted in this section.

4.2 Proposed Obstacle Avoidance Strategy

A fundamental function of the wearable assistive device for VIPs is its ability to avoid obstacles. Without an effective obstacle avoidance strategy, ensuring the user's safety becomes challenging, with a higher likelihood of collisions and hindered mobility. Navigating through environments filled with potential hazards and obstacles poses significant challenges for the visually impaired. Achieving effective obstacle avoidance is a complex task, requiring algorithms that consider various features, including physical characteristics, location, height, distance, and pathway blockage in the information provided to users. While several obstacle recognition and avoidance algorithms have been proposed, many fall short in considering these features in the provided information In our strategy, we leverage a stereo camera carried by the VIP to capture scenes in front of the user and a custom lightweight YOLOv5 to detect objects in the scene. We prioritize objects that are closest and obstructing the user's path as potential obstacles. Our assumption is that objects near the user and along the user's path are most likely to be obstacles. By prioritizing the locations of close obstacles, we aim to identify and distinguish between areas that are dangerous and safe for navigation. Areas that are considered are entrances, intersections, and potential hazards and obstacles in the path of the VIPs. Simultaneously, the system focuses on detecting objects in close proximity to the user, providing immediate feedback and guidance to avoid collisions or potential dangers. As depicted in Figure 4.1, the Image $I(x, y)$ has an adaptable grid represented with a yellow color, and the aqua area is obstacle to be avoided by the user by moving through the shade region.

In Figure 4.1, we illustrate how the system checks for obstacles using the adaptable grid. The grid checks the location of the detected objects by checking if the cells

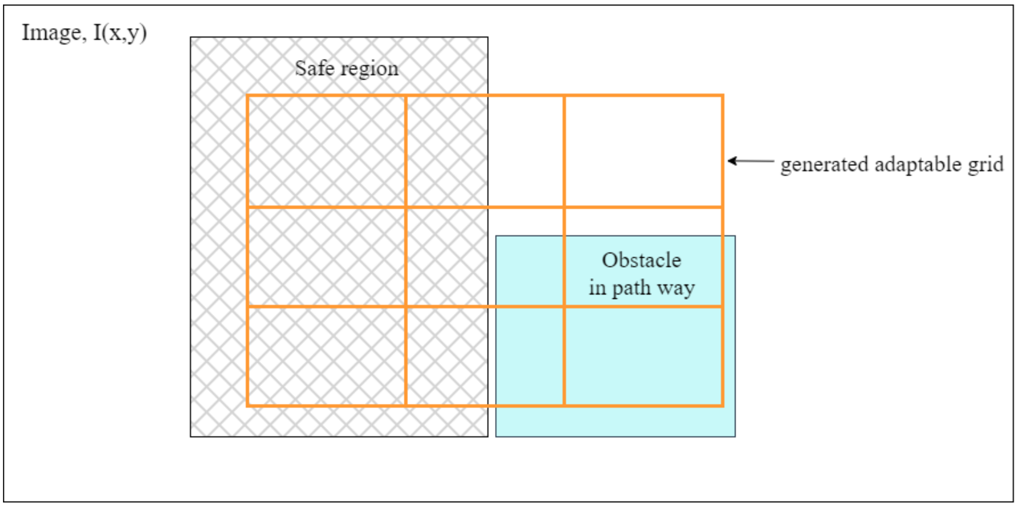


FIGURE 4.1: Illustration of the adaptable grid to check the safe region for the user to traverse.

of the grid overlap with the object. The region of the cells with no overlaps becomes the safe region the user can traverse. The

To implement this, we establish two sets of conditions for identifying obstacles among detected objects using our object detection models:

- Objects must be within a 2-meter threshold distance to be considered obstacles, with priority given to the closest objects.
- central area of the captured image scene is designated as the VIP's pathway, and objects within this region are recognized as potential obstacles.

This strategy ensures a comprehensive evaluation of scene objects, both collectively and individually, to select the most immediate obstacle that the VIP should avoid. This is because the approach checks each detected objects in fast and real-time and dynamically assign the status to them before the selection of the closest and most hazardous objects is selected. The implementation of this strategy is presented in a in Figure 4.2

4.2.1 Object Detection

Object detection is the task of locating and classifying objects within an image. Typically, this involves drawing bounding boxes around the objects in the image and assigning a label to each box that describes what the object is (e.g., "car", "person", "tree", etc.). Object detection such as single-step detection algorithms have improved tremendously in speed and accuracy making them useful in diverse applications such as autonomous driving, robotics, and surveillance. Among the single-stage detectors, we choose YOLO for being extremely fast and accurate even on microcontrollers.

These speed and accuracy makes YOLO method a good candidate for real-time applications. YOLOv5 is the fifth version in a series improvement on the YOLO architecture, and it comes with several improvements over its predecessors. One of the standout features of YOLOv5 is its balance between accuracy and speed. It achieves high accuracy while being efficient enough to run in real-time on modern hardware. The YOLOv5 architecture consists of multiple detection layers, including

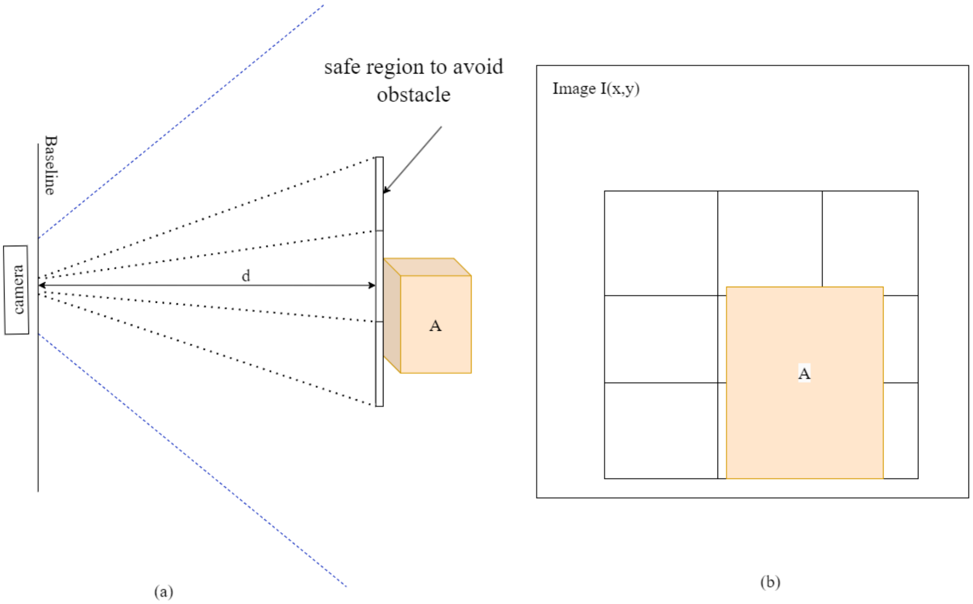


FIGURE 4.2: Illustration of the overview of adaptable grid to select an obstacle based on our assumption conditions of a critical distance and critical region to avoid obstacle

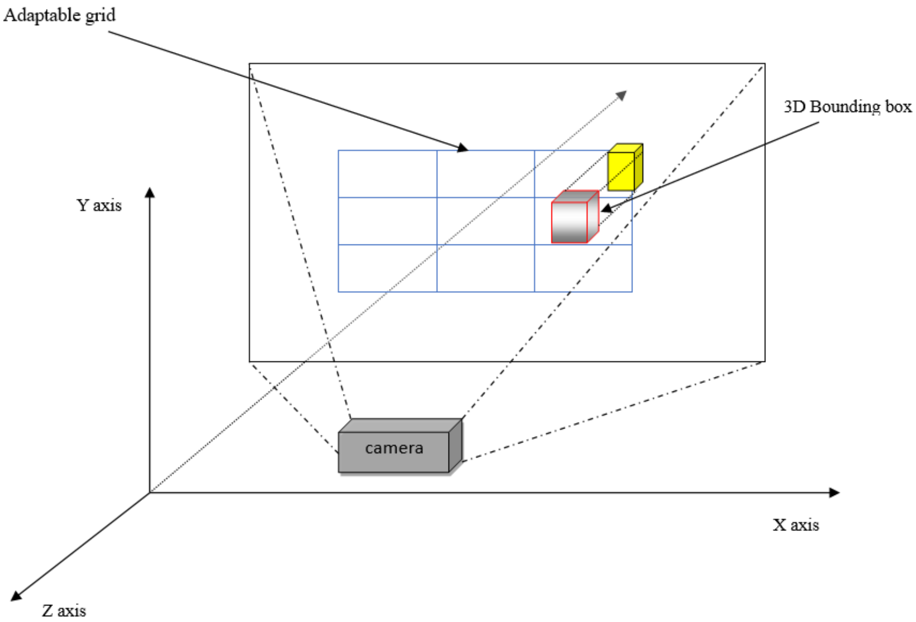


FIGURE 4.3: An illustration of the system that demonstrates that relates the camera to the 3D bounding box marked with red edges and an obstacle depicted with a yellow box in the location of the adaptable grid system for avoiding the obstacle. We consider the camera coordinate system and world coordinate system to determine how the user is supposed to move around obstacles based on the feedback of free spaces on the adaptable grid.

S, M, L, and X, each optimized for different object scales. For a smaller size and memory efficiency with decent use of computational resources on microcontroller, we choose YOLOv5s model as our pretrained models to fine tunning for the object detection.

YOLOv5 architecture comprises the following components

- Backbone: YOLOv5 employs a CSPDarknet53 as its backbone network. This architecture introduces Cross Stage Partial connections, reducing computational overhead while preserving accuracy.
- Neck: PANet (Path Aggregation Network) is used as the neck architecture. PANet enhances feature fusion across multiple scales, allowing the model to capture objects of various sizes effectively
- Head: YOLOv5 uses a detection head comprising multiple detection scales (S, M, L, and X) to detect objects at different sizes and resolutions. Each scale predicts bounding boxes, class probabilities, and objectness scores.

For our dataset, we selected OID (Open Images Dataset) [77] which is a large-scale dataset containing millions of labeled images across a wide range of object categories. While it is a valuable resource for training object detection models, it may not always cover specific domains or use cases adequately.so we adapted it by filing in with data on objects which could possible be obstacle but not part of the OID dataset involves selecting a subset of relevant images and labels, often focusing on a specific task or domain which is the detection of objects that can pose as obstacles.

Training a YOLOv5 model involves fine-tuning the pre-trained weights on your customized dataset. The training process typically consists of multiple epochs, where the model learns to detect objects in your specific domain. It's essential to monitor training metrics such as loss and mAP (mean Average Precision) to assess the model's performance.

YOLOv5 provides several hyperparameters that can be adjusted to improve model performance. These include learning rate, batch size, and anchor box configurations. Hyperparameter tuning often involves experimentation to find the optimal values for your dataset.

The customized YOLOv5 model showed promising results in terms of both accuracy and speed. Its lightweight architecture made it suitable for deployment on our resource-constrained NVIDIA Jetson Nano device. In real-world testing scenarios, the model demonstrated the ability to accurately detect and classify obstacles in near real-time, making it an excellent choice for autonomous navigation tasks. We present a confusion matrix, a precision and recall graphs of the training results. The confusion matrix in Figure 6 shows very good results for accurately predicting the right objects within the images. The plot of the precision shown in Figure 7 and the plot of the recall show in Figure 8 show that the model could precisely detect the objects it was trained on. The experiments were conducted to assess the accuracy of object detection using YOLOv5, as well as the classification of objects and obstacles based on their distance and their placement within the cells of the adaptable grid.

4.2.2 Depth Estimation

Depth estimation using stereo cameras is a technique that attempts to infer the distance of objects in a scene from the difference in their appearance between two images taken from slightly different viewpoints. This concept is analogous to human stereoscopic vision, where the disparity in the position of an object as perceived by our left and right eyes enables our brains to gauge depth. In a stereo camera setup, two cameras are placed parallel to each other at a known distance apart, called the "baseline." The two cameras simultaneously capture images of the same scene. The difference in the position of an object in the left and right images is called disparity.

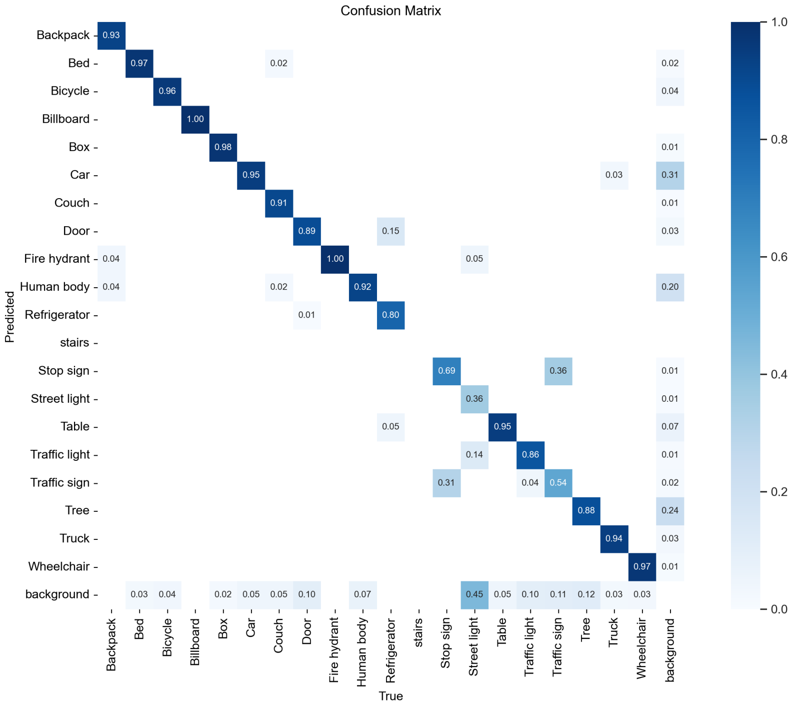


FIGURE 4.4: A confusion matrix of the model YOLOv5 predictions during training. The confusion matrix is a representation of performance predictions of the model each class,

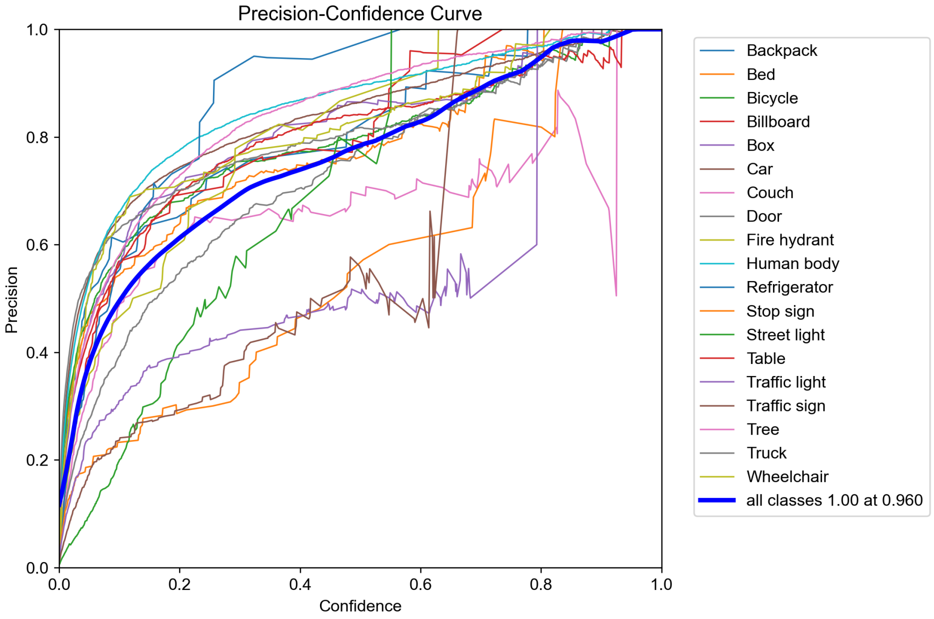


FIGURE 4.5: A confusion matrix of the model YOLOv5 predictions during training. The confusion matrix is a representation of performance predictions of the model for each class,

Objects closer to the cameras will have a higher disparity, while those farther away will have a lower disparity. With the known baseline distance and the calculated disparity, you can use trigonometry to estimate the distance (or depth) to the object. This method of determining depth is called triangulation.

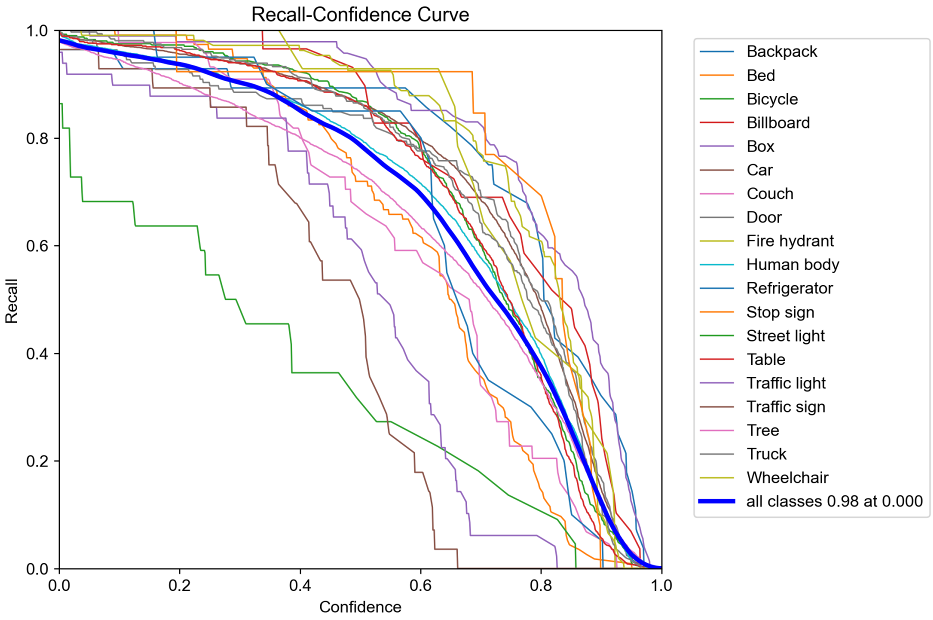


FIGURE 4.6: A confusion matrix of the model YOLOv5 predictions during training. The confusion matrix is a representation of performance predictions of the model for each class,

$$Depth = \frac{f \times b}{disparity} \tag{4.1}$$

where f is the focal length of the cameras, and b is the baseline distance between the cameras. This $disparity$ is a map representation of the scene's depth, where each pixel's value indicates the disparity of the corresponding point in the scene.

4.2.3 Obstacle Detection Approach (Adaptable Grid System)

We explain how the adaptable grid is constructed and use in obstacle avoidance in this section. After detecting objects in the scene with their 3D bounding boxes being determined, we generate a 3x3 grid dynamically from the center point of the image for every object to determine their location with respect to the path of the user of the devices. This helps us to determine if the objects fall into the path of the user. The apparent size of the object in the scene makes it difficult to adapt to this Figure 4.7 shows the ada

In our proposed obstacle detection approach which is based on object detection, we classify detected objects whether it is an obstacles or not for visually impaired persons. The innovation in this approach is that it runs fast and in real-time since it simultaneously checks all the detected objects at once.

The adaptable grid helps us to determine regions that are safe for the user to traverse. This adaptable grid approach is grounded in two key assumptions: the importance of prioritizing locations and the proximity of obstacles to VIPs. We recognize that not all areas within the environment carry the same level of significance for navigation. Therefore, to formulate our strategy, we assume that having the camera strapped to the chest while walking in the direction of the camera's field of view is the most reliable position for capturing any obstacles, whether static or dynamic, in front of the VIP. We also assume that the path taken by VIPs will pass through the central point of the image captured by the strapped camera. Consequently, any object in this proximity is likely to pose a collision risk to VIPs. Moreover, objects closer to the user are considered more hazardous. These assumptions and descriptions are detailed in Figure 4.2.

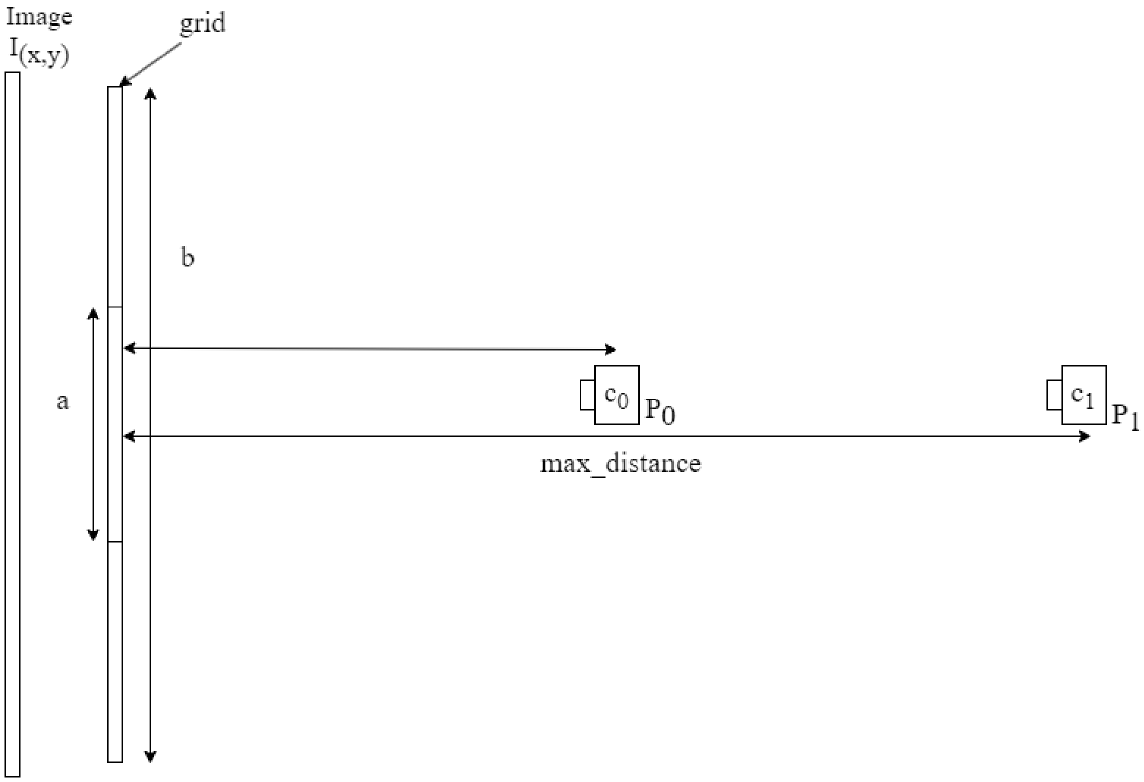


FIGURE 4.7: An illustration of various distance and camera configurations for creating the dynamic grid for selection of the obstacles. The dynamic grid size generation is based on the linear relationship between the distance and the size of the grid to be generated

The distance from the camera to the observed object is measured in meters. To make the grid size adaptable to the apparent size of the object in the image, we need to normalize this distance to a range between 0 and 1 using a min-max scaling approach. We divide the actual distance by *max_distance* to obtain a normalized distance value between 0 and 1. This normalized distance represents how close the object is relative to the maximum distance. The normalization is given by equation 4.2 The adaptable grid size helps not only in the selection of the obstacle but determining the apparent size of the objects so we can be aware of the impact it we can adjust it. This helps to guide the blind successfully

$$normalize\_distance = \min(\frac{distance}{max\_distance}, 1.0) \tag{4.2}$$

To determine the size of the adaptable grid for each object based on their normalized distance, we conducted visual observation to empirically establish the minimum and maximum grid sizes that can bound a detected object as *max_grid_size* and *min_grid_size* in pixels within the frames captured respectively. The calculation for the grid size is defined by the following equation 4.3

$$grid\_size = min\_grid\_size + (max\_grid\_size - min\_grid\_size) \times (1 - normalized\_distance) \tag{4.3}$$

The *max_grid_size* is the maximum size the grid can have in pixels when objects are very close to the camera. While *min_grid_size* is the smallest size the grid will have when objects are at or beyond the *max_distance*. As illustrated in Figure 4.7, the grid will have the smallest size, a for a detected object when the camera is at a point P1 and a distance, c_1 distance *max_distance* and while the grid will have a size, b, almost covering the whole image when the camera is at point, P_0 and a distance, c_0. This configuration ensures that the grid adjusts well to the region along the pathway covered by the obstacle

4.3 Experiments

We carried out an experiment to evaluate the performance of our assistive device's obstacle avoidance model. The evaluation focuses on the measurement of the distance between the camera of the user and the observed obstacles, the accuracy of the objects, the selection of the obstacle among the detected objects, and the choosing of one of the obstacles that poses an imminent danger to the user of the system.

4.3.1 Experimental Objectives

The objective objective of the experiment is to evaluate performance and robustness of accuracy in detecting and selecting the obstacles, response time, and adaptability to the enviromental conditions. We set the following primary objectives of this experimental design are as follows:

- To evaluate the accuracy and effectiveness of obstacle detection and avoidance in diverse environmental conditions.
- To measure the system's response time in providing alerts and guidance to users.
- To investigate the impact of environmental factors, such as lighting conditions and obstacle types, on system performance.

4.3.2 Experimental Setup

We outline the experimental objectives and setups. The experiment's goal is to assess the device's performance in detecting potential obstacles and selecting the most immediate threats to the user's mobility.

This experiment aims at a comprehensive approach to evaluating an obstacle avoidance system for the visually impaired. By combining quantitative and qualitative measures and considering diverse environmental factors, the study aims to provide valuable insights into system effectiveness and user satisfaction. The results of this experiment have the potential to inform the development of a more robust obstacle avoidance strategy into assistive technologies for the visually impaired.

4.4 Results and Discussion

In this section, we present the results of testing the avoidance strategy in an indoor environment and discuss their implications. We analyze the accuracy of object classification as obstacles, distance measurements, and the total execution time for the obstacle avoidance strategy.

To calculate the precision with which the systems classify objects and obstacles, we utilize the following equations to assess precision (sensitivity of obstacle classification, presented in equation 4.4), recall (specificity of obstacle classification, presented in equation 4.5), and accuracy of object selection as obstacles in equation 4.6.

$$Precision : \frac{TP}{(TP+FN)} \tag{4.4}$$

$$Recall : \frac{TN}{(TN+FP)} \tag{4.5}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.6}$$

TABLE 4.1: A confusion matrix on the accuracy of the recognition of obstacles from the list of objects detected by the custom YOLOv5

		Predicted Value	
		Detected as Objects	Detected as Obstacle
Actual Values	Detected as Objects	95%	5%
	Detected as Obstacle	4%	96%

Table 4.1 shows a confusion matrix illustrating the accuracy of the detected obstacles based on the selection the strategy we presented. The results demonstrate an impressive accuracy of 95%, with only a few false positives recorded.

TABLE 4.2: A confusion matrix on the accuracy selection of most hazardous obstacles from the set of obstacles based close proximity

		Predicted Value	
		Critical Obstacle	Selected Obstacles
Actual Values	Critical Obstacle	93%	7%
	Selected Obstacles	6%	94%

Table 4.2 presents another confusion matrix, this time focusing on the accuracy of our algorithm to select the most hazardous among the detected ones. The results indicate that our approach can effectively select the right obstacle and promptly inform the user

As usual, the diagonal elements represent the accurately predicted objects and obstacles in the experimentation. Table 3 shows how well the strategy selects all obstacles from the detected objects, while Table 4 focuses on choosing the immediate obstacle, critical obstacle, to inform the user of the hazard in front of them

Many of the systems focus on recognizing or detecting obstacles but lack specificity regarding which obstacle poses the highest risk to the user. While a few mention obstacle selection, the criteria for prioritization and informing the visually impaired are often

TABLE 4.3: Distance ranges of the objects detected from the camera and the adaptable grid size for the objects

Obstacles	Actual distance, AD (m)	Predicted distance ,PD (m)	Error (AD-PD) (%)	Grid size
chair	0.90	0.95	5.50	(737x737)
Table	1.50	1.48	1.33	(695x695)
Persons	1.80	1.75	2.77	(694x694)
chair	1.05	1.05	0.00	(726x726)
Persons	1.80	1.75	2.77	(695x695)
Fridge	1.15	1.20	4.34	(720x720)

TABLE 4.4: Detecting obstacles in the path with the shortest distance and in the critical region prioritized

Condition	Objects	Obstacles	Shortest Distance	Prioritized?	Processing Time(s)
Normal	5 out of 6	2 out of 3	0.9	Yes	0.25
Bad lighting	7 out of 7	4 out of 5	1.5	Yes	0.4
Cluttered Area	15 out of 15	3 out of 4	0.7	Yes	0.5
Unstable Camera	8 out of 10	4 out of 5	0.8	Yes	0.31

In addressing this gap, we considered a scenario with multiple obstacles, emphasizing how the obstacle detection system chooses which obstacle presents the most risk and how to navigate around it. The selection criteria which are based on the assumption of prioritizing the obstacles closest to and directly in harm's way of the user ensure that the user is alerted and informed on how to evade it. Most related works did not explicitly address this issue. Our selection process is based on the assumptions outlined in Section 4.2.3.

The qualitative evaluation of most wearable assistive devices was calculated using a formula presented in Tapu *et al.* called the Global Score:

$$GlobalScore = \frac{\sum_{i=1}^{N} w_i \times F_i}{N} \tag{4.7}$$

where F_i is the score assigned to the ith feature, N is the number of characteristics used in the evaluation, and w_i is the weight assigned to each feature. The features considered in the qualitative evaluation include processing speed, usability, robustness, coverage distance, obstacle detection, portability, and friendliness. This set of features has become a standard for comparison in many research reviews. Below are the definitions of the features considered in the evaluation:

- Processing speed: The device should operate in real-time, and feedback should be timely for the user's response to obstacles at a minimal distance of 1.5 m.
- Usability: The device should function in both indoor and outdoor environments.
- Robustness: The system should not be influenced by scene dynamics or lighting conditions
- Coverage distance: The maximum distance between the user and the object should be considered so that the system can detect the object.
- Obstacle detection: The system should be able to detect any object regardless of the shape, size, and state of the object.
- Portability: The device should be ergonomically convenient to wear and move with.

- Friendliness: The device should be easy to operate.

Our assumptions emphasize that proximity obstacles pose the greatest risks and that the adaptable grid, flexible in adapting to the apparent shapes and sizes of images, leads to a more accurate approach in selecting the most imminent obstacle to avoid. We believe that identifying and determining the types of obstacles and the risks they pose to VIPs are crucial for ensuring their safety while navigating in both indoor and outdoor environments.

TABLE 4.5: Comparison of our proposed obstacle avoidance strategy with related research in terms of features.

Features	**Mechatronic[78]**	**Moldanu et al. [79]**	**Jafri et al. [80]**	**Sound of Vision [81]**	**Everding et al. [82]**	**Shwarze et al. [83]**	**Ours**
Type	Monocular-Based	Monocular-Based	RGB-D-Based	RGB-D-Based	Stereo-Based	Stereo-Based	Stereo-Based
Usability	Outdoor	Indoor/ Outdoor	Indoor	Indoor/ Outdoor	Indoor	Outdoor	Indoor/ Outdoor
Coverage distance (m)	10	5	2	5–10	6	10	10–20
Shape and size	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Portability	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obstacle detection	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prioritization	No	No	-	No	No	Yes	Yes
Accuracy selection	-	-	-	-	-	-	93%
Score [33]	5.86	8.74	5.69	8.19	8	8.32	8.80

4.5 Conclusion

In this study, we introduced an efficient obstacle avoidance method utilizing an adaptable grid-based strategy. Our goal was to create an approach capable of not only real-time obstacle detection but also dynamic flexibility, adjusting its grid structure to enhance obstacle avoidance accuracy across different settings and scenarios. We designed a wearable assistive device known as ETA, primarily to assist visually impaired individuals in identifying and evading obstacles. Our obstacle detection system exhibited an impressive detection rate, reaching as high as 95% in specific environmental conditions. The system effectively delivers timely and pertinent information to visually impaired individuals, empowering them to navigate safely and independently in diverse surroundings. By addressing the distinct challenges encountered by VIPs, our research contributes to enhancing their quality of life and promoting their active participation in society. Looking ahead, we aim to extend this approach to accommodate non-rigid objects and various types of ground surfaces, further refining its capabilities for the future.

Chapter 5

Speech Communication Interface with Request Support

5.1 Introduction

In an era marked by rapid technological advancement, voice-controlled applications are gaining widespread popularity due to their convenience and accessibility. This component of our research explores the importance of speech communication techniques as voice assistants for VIPs. The voice assistant systems are often deployed on micro-controller-based for command or request support. These systems are used in various applications for different tasks, such as customer support services. Speech communication modules are often made up of speech recognition, speech enhancement, and speech synthesis technologies into a single system that can be used to support VIPs in their navigation.

The convergence of these technologies has the potential to greatly improve accessibility to navigation information and independence for the blind during mobility, enabling them to interact with assistive devices and access information more effectively. For example, the user can ask for the routes to their destination and what their environment looks like. Speech recognition and synthesis systems have become integral components of human-computer interaction. They offer a natural and intuitive way for users to communicate with electronic devices. In this command request application where users issue verbal instructions to request some action in the scope of the functionality of the assistive navigation systems, integrating speech technology can enhance user experience and accessibility in terms of feedback of alerts and responses.

Our speech communication system comprises these three components for interfacing the communication between the user and the assistive navigation devices as shown in 5.1. The three components are the speech recognition module for recognizing words in the speech of the user, the speech enhancement module for removing the background noise in the speech making it clear and easy to recognize the words, and the text-to-speech module for outputting speech when provided text from the other systems. The module is supposed to support the alert systems from the other modules and respond to the user. The module has a Speech recognition function, the module is a cornerstone of this integrated system. It allows users to interact with the system through spoken commands. For the blind, this technology can serve as an intuitive means of controlling devices and accessing information. The models have been the building blocks on which several other research tasks have been promoted.

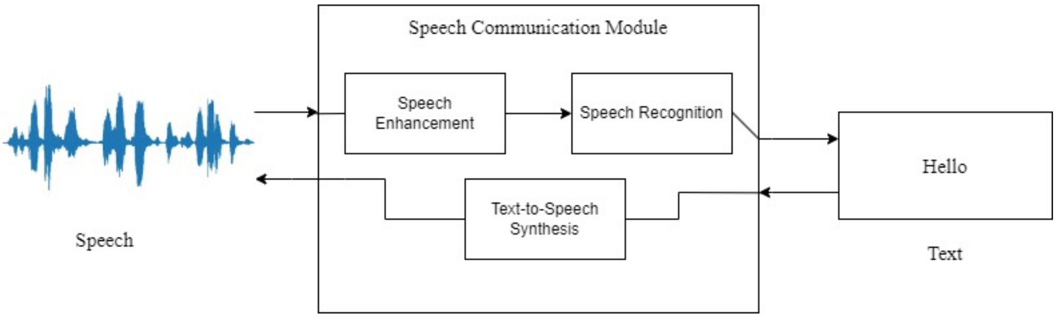


FIGURE 5.1: Illustration of the three components in a speech communication module for voice assistant systems and the role they play.

5.1.1 Speech Recognition

Speech recognition, also known as Automatic Speech Recognition (ASR), is the process of converting spoken language into text or commands for the system to work on. For a microcontroller-based command request module, a speech recognition system can be implemented using various techniques, including Hidden Markov Models (HMMs) [84], Deep Neural Networks (DNNs) [85], or more advanced models like Deep Convolutional Networks (DCNs) [86] or Transformers [87].

Speech recognition is the fundamental component of voice-controlled command request systems. The primary objective is to convert spoken words or phrases into machine-readable data that can trigger specific actions. On resource-constrained microcontrollers, achieving accurate and real-time speech recognition poses several challenges.

To ensure the accuracy and robustness of speech recognition, speech enhancement features play a crucial role. These features involve noise reduction, echo cancellation, and signal processing techniques to improve the quality of incoming speech signals. Noise and environmental factors can significantly affect speech recognition performance, especially in real-world applications. Therefore, the integration of speech enhancement features is essential for reliable operation.

5.2 Speech Enhancement

Assistive devices for the blind often rely on speech synthesis and recognition technologies to convey information to the user and receive input from them. For example, in a system where audio feedback mechanisms are implemented, the alert cues the VIP need are presented as a synthesized speech to them. In the situation where the VIP wants to request assistance, a speech recognition module needs to recognize and translate the speech for the device to operate on. In crowded or busy urban environments, the background noise could distort the speech in Speech enhancement ensuring that the synthesized speech is clear, natural, and easily understandable, facilitating effective communication between the device and the user.

To address this, we research possible ways to provide an effective background noise suppression model for our assistive devices. We considered the following state-of-the-art models for speech enhancement and designed an improved version [88]. We address the issue for background using a Generative Adversarial Network, (GAN) approach introduced by Goodfellow *et. al* [89] in which a generator is

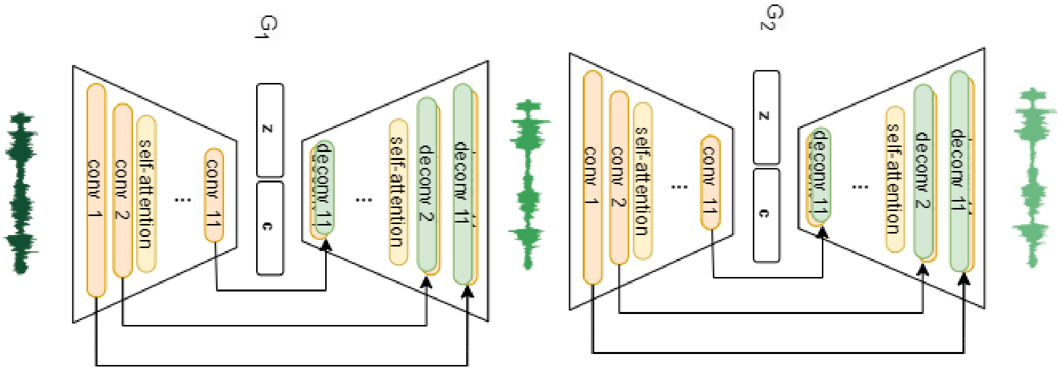


FIGURE 5.2: Illustration of our proposed multi-stage GAN for the speech input to the microcontroller device to enhance the input speech before recognition. The model takes a noisy signal represented in deep green passed through generators G_1 and G_2 to output a clean speech represented with a light green.

trained adversarially with a discriminator such that the generator tries to synthesize output plausible outputs identical to the targets and the discriminator makes efforts to distinguish the fake from the real in a min-max sum zero settings. Based on the GAN method, Mirza and Osindero proposed a conditional GAN [90] and used to translate the image-to-image task [91]. This breakthrough led to the seminal work on GAN-based speech enhancement by Santiago *et. al.* [92] and several modifications as well [93]–[95]. Later the multi-generator GAN was used in speech enhancement by Phan *et. al.* [96] to clean speech using a multistage approach in which two (2) or generators are trained adversarially with a discriminator such that the generator tries to synthesize the output. Due to the number of convolutional steps in the multistage enhancement, some of the temporal dependence is lost in the process [94]. We explore adding layers of self-attention to the multistage process of speech enhancement with a number of generators which yielded a better result.Figure 5.2 show one of the setup from our experiment using two (2) generators with self attetion layers to generate enhanced speech.

TABLE 5.1: The results of the ablation study on $N = 2$ (i.e., 2 generators) setup of the ISEGAN-Self-Attention and the DSEGAN-Self-Attention networks. The results are compared with ISEGAN and DSEGAN. The **boldface** is the setup with the highest score per an objective evaluation metric.

Metric	**DSEGAN**	**ISEGAN**	**ISEGAN-Self-Attention**		**DSEGAN-Self-Attention**	
			G_1	G_2	G_1	G_2
PESQ	2.71	2.66	**2.63**	2.57	**2.68**	2.64
CSIG	3.58	3.58	**3.51**	3.49	**3.52**	3.50
CBAK	3.15	3.15	**3.09**	3.08	**3.11**	3.08
COVL	3.11	3.04	**3.07**	3.04	**3.09**	3.05
SSNR	9.19	9.04	**9.11**	9.03	**9.09**	9.02
STOI	93.29	93.25	**93.38**	93.32	**93.42**	93.36

We present the results in Table 5.1 where we compared our results with state-of-the art methods such SEGAN, SASEGAN, ISEGAN and DSEGAN. The introduction

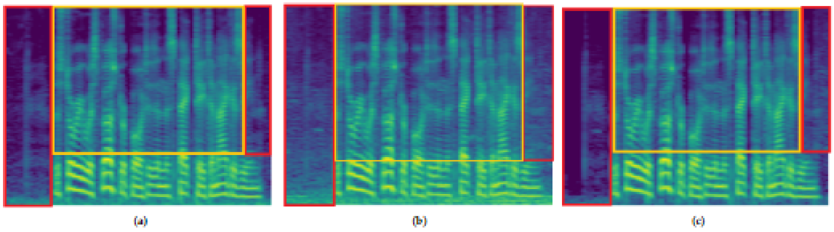


FIGURE 5.3: Spectrograms of a speech signal generated using Librosa [97]: (a) the target clean speech signals, (b) the noisy speech signals, and (c) the enhanced speech signals with the different networks. The enhanced speech signal produced by our model. The regions marked red are regions with high noise levels with little or no intelligible speech signal, and the regions marked with yellow rectangles contain intelligible speech signals with little additive noise. It can be seen in (c) that the networks can remove additive noise in the red rectangles as well as in the regions marked with yellow rectangles. The spectrograms in (c) show that the speech signals are preserved while the networks remove the noise.

of the self attention saw a significant improvement in the enhanced speech measured using speech measurement metrics. The first and second best results for each metric is boldened to distinguish them from the rest of the results. The results shows that adding the self attention layer to multigenerator with shared parameters will yield a better by approximately 10

Figure 5.2 shows the spectrogram images of the result of our model for the speech enhancement model. In (a) the ground truth or target spectrogram, (b) the noisy spectrogram we are trying to clean, and (c) the cleanse spectrogram. We marked the regions where most of the background noise has been removed.

5.2.1 Speech Synthesis

Speech synthesis, also known as Text-to-Speech (TTS) conversion, is the process of generating human-like speech from text input. In the context of a microcontroller for command request applications, TTS enables the system to provide alerts during encounters with hazardous situations and respond verbally to user commands or provide feedback. TTS engines can be rule-based like [98] or use machine learning techniques like WaveNet by Oord *et. al.* [99] or Tacotron by Wang *et. al* [100] for more natural-sounding speech At a higher level, most of the speech synthesis converts sentences or a set of input characters into speech or audio signal. The process involves generating spectrograms for the given characters and then constructing them into waveforms. With our system, the modeled approach for the deep learning model is similar to speech Deep Convolutional Network by Tachibana *et. al* [101]. The modifications made to this model are to ensure that it runs faster with minimum resources for prediction.

The speech synthesis model which is responsible for converting the text to speech, was tested using the Mean Opinion Score. This is standard metric used in the Telecommunication industry to check the quality of speech information transmitted from one point to another.

In determining the Mean Opinion score we conducted a survey. 10 participants were selected to listen to the speech generated by our model and measure the quality using the specific criteria for MOS. The data analyzed and used to determine the

MOS score for the two datasets that was used to train the models for speech synthesis.

TABLE 5.2: Subjective 5-scale MOS in naturalness and Intelligibility performance score for our text-to-speech model

TTS Model	MOS Score
DC+SSRN	3.56 ± 0.093

5.3 Speech Communication Module

Having discussed the various components we worked on for the speech communication module. We present the model for providing communication between the user and the assistive device. Figure 5.4 depicts the model where the user provides an input speech command the speech recognition function converts it to a text which is worked on by the system as request and generate a respond for the request. The system also use the speech synthesis module to convert the text respond into a speech which the VIP can asses and listen to. This way there can be a medium of commmunication between the device and the user.

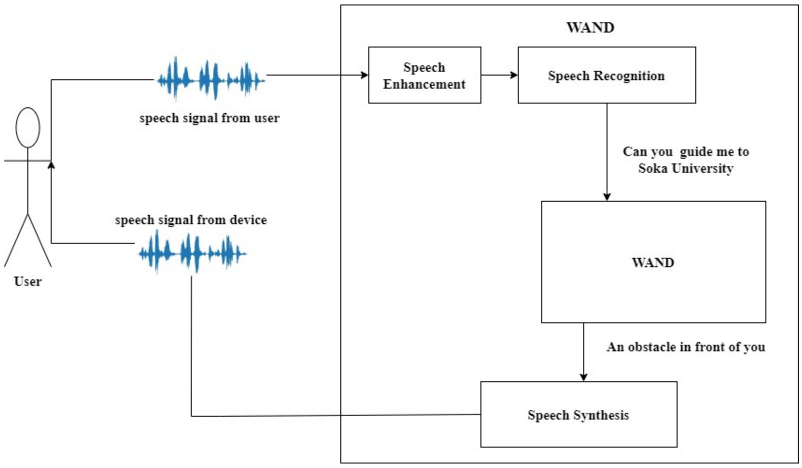

WAND
Speech Enhancement
Speech Recognition
speech signal from user
Can you guide me to Soka University
speech signal from device
WAND
User
An obstacle in front of you
Speech Synthesis

FIGURE 5.4: An illustration of the speech communication module for supporting the visually impaired showing the various components and flow of information

To demonstrate the flow of the speech data and text between the user and the device, an illustration in Figure 5.5 to show the timelines of the processing and delivery of the diferent modal data to the device or the user.

5.3.1 Implementation of the Speech Communication Module

Our device is a microcontroller with low computational resources and it is expected to be fast for real-time application of alerting and processing the request of the user. Therefore, we rely on application programmable interfaces APIs to run the modules in the cloud and deliver the results over the internet rather than performing speech recognition, speech synthesis, and speech enhancement directly on the device.

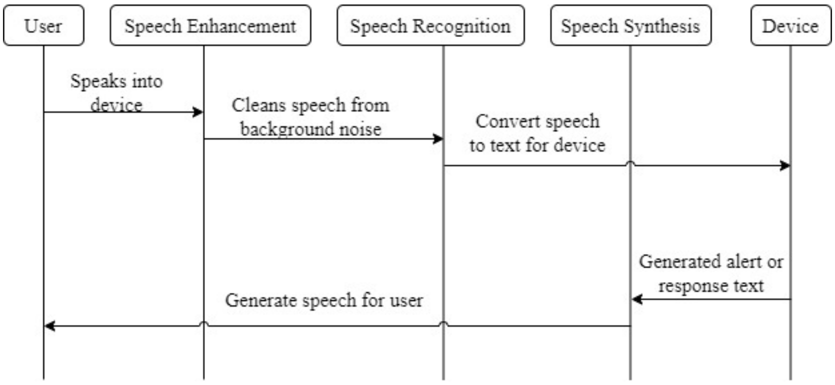


FIGURE 5.5: An illustration of the speech communication module for supporting the visually impaired showing the various components and flow of information

Microcontrollers typically face limitations in processing power, making it challenging to directly run complex speech recognition algorithms. To circumvent this, we employ a strategy where speech recognition, speech enhancement and speech synthesis tasks are offloaded to cloud-based servers. Once processed, the results are delivered back to the microcontrollers through APIs.

Given the memory constraints inherent to microcontrollers, this approach minimizes the need for large-scale data storage and complex model management on the device itself. Techniques like quantization and pruning, which are commonly used to streamline recognition models for local execution, are less critical in this setup since the heavy computational work is handled in the cloud.

This cloud-based processing is particularly beneficial in real-world environments, where noise and variability in spoken commands can significantly challenge recognition accuracy. By leveraging robust speech recognition algorithms and advanced noise cancellation techniques available on the server, we ensure more reliable performance. The cloud server processes the audio input, filters out noise, and interprets the commands, sending concise, actionable data back to the microcontroller. This setup allows the microcontroller to effectively alert and respond to the needs of the user, despite its limited local processing capabilities.

5.4 Conclusion

The integration of speech recognition, speech enhancement, and speech synthesis technologies into a microcontroller-based system offers tremendous potential for assisting the blind in their daily lives. Even though we were able to achieve the goal of building the models to run offline and on the micro device, the low computational capability of the microcontroller makes it difficult to use it in real-time therefore our reliance on the online APIs. We hope in the future we can speed up the models and make them work in real-time. This comprehensive approach empowers users to interact with devices, access information, and control their environment through intuitive spoken commands. As technology continues to advance, the future holds even greater promise for improving accessibility and enhancing the quality of life for individuals with visual impairments.

Chapter 6

Conclusion

6.1 Introduction

In the fast-paced and ever-evolving landscape of technology, the development of wearable assistive navigation devices with obstacle avoidance, pathfinding, and plane detection functionalities represents a remarkable leap forward in improving the quality of life for individuals with mobility challenges. This concluding chapter reflects on the journey of creating and implementing such devices, highlighting their significance, potential impact, and areas for future exploration.

6.2 Summary of Our Wearable Assistive Navigation Devices

We introduced a wearable assistive navigation device called WAND, consisting of several key components: a pathfinding module with tracking functionality, an obstacle avoidance module, a ground plane detector module, and a command support module. This device, designed as an Electronic Travel Aid (ETA), was created to assist visually impaired individuals in detecting and avoiding obstacles, ensuring safe navigation, and providing guidance.

We discussed an effective obstacle avoidance strategy based on an adaptable grid approach. The primary goal was to develop a strategy capable of real-time obstacle detection while dynamically adjusting its grid representation to enhance obstacle avoidance accuracy across different environments and scenarios. Notably, the obstacle avoidance module exhibited a high detection rate of up to 95% in specific environments and 93% for high-risk obstacles.

The pathfinding with tracking module utilizes data from OpenStreetMap to identify coordinates lying between the user's current location and their destination, creating a safe path for navigation. The tracking feature ensures that the user's current coordinates match or are near their actual location. If a deviation from the path occurs, the user is immediately alerted.

The ground plane detector module plays a crucial role in identifying safe floor surfaces, enhancing the user's safety during navigation. The module uses a point cloud to create a mesh and determine if the mesh plane is horizontal for the user to traverse if the mesh plane is not horizontal the VIP is informed about the unlevelled ground plane in front of him.

Additionally, the command support module facilitates speech communication between the user and the device. It captures the user's spoken commands, converts them into text-based instructions for the device to execute, and provides alerts and information to the user through speech output.

Overall, this system offers visually impaired individuals timely and relevant information, empowering them to navigate various environments independently and safely. By addressing the unique challenges faced by VIPs, our research contributes to improving their quality of life and promoting their inclusion in society. As part of future improvements, we aim to adapt this approach to handle non-rigid objects and diverse ground surfaces.

6.3 Significance of our Wearable Assistive Navigation Devices

The device provides an opportunity for people with visual impairments to regain a sense of independence and freedom that may have previously been limited. Integrating obstacle avoidance, pathfinding, and plane detection functions into a single wearable device has the potential to support VIPs around the world with navigation. The solution can also be considered an effort towards achieving SDG goals number one (1), which is fighting poverty, three (3), which relates to good health and well-being, and ten (10), aimed at reducing inequalities.

6.4 Impact on Mobility and Quality of Life

The primary goal of our proposed navigation device is to enhance mobility and, consequently, the overall quality of life for users. These devices enable users to navigate their environments with greater ease and confidence, reducing the barriers they face in everyday activities. From crossing busy streets to moving through unfamiliar indoor spaces, these devices provide users with invaluable assistance, thereby fostering a more inclusive and accessible society. We hope that in the near future, the device can undergo all experimental testing, pass qualitative testing, and become affordable to VIPs all around the world.

6.5 Technological Advancements and Challenges

The development of wearable assistive devices represents a fusion of cutting-edge technologies. Machine learning, deep learning algorithms, and computer vision have converged to create devices that can not only detect obstacles and plan optimal routes but also identify various types of surfaces and changes in terrain. However, this remarkable progress is not without its challenges. Battery life, computational power, and device ergonomics are just a few areas that require ongoing improvement. A critical aspect of designing wearable assistive navigation devices is ensuring that they are user-centered and inclusive. User feedback and iterative design processes have played a pivotal role in shaping these devices to meet the diverse needs of their users. The inclusion of individuals with disabilities in the design and testing phases is essential to creating devices that are truly effective and user-friendly.

6.6 Future Directions

In closing, wearable assistive navigation devices with obstacle avoidance, pathfinding, and plane detection functionalities are emblematic of the remarkable possibilities that technology can offer to enhance the lives of individuals with disabilities. These devices exemplify the intersection of innovation, compassion, and inclusivity. As we move forward, let us remain steadfast in our dedication to pushing the boundaries of what is possible, ensuring that these devices reach those who need them most, and continue to make strides in creating a more accessible and inclusive world for all.

As we conclude this journey in the development of wearable assistive navigation devices which is just the begining to greater innovation for assistive device developments, it is essential to consider the exciting avenues that lie ahead. The ongoing advancement of technology promises even more sophisticated functionalities, increased reliability, and greater affordability. Additionally, integrating these devices seamlessly into existing infrastructure and services will be crucial in realizing their full potential.

Furthermore, collaborations between researchers, developers, healthcare professionals, and individuals with disabilities will continue to drive innovation in this field. The open exchange of ideas and the collective commitment to improving the lives of those who rely on these devices will fuel further progress.

6.7 Conclusion

In conclusion, the development of wearable assistive navigation devices with obstacle avoidance, path finding, and plane detection functionalities represents a significant leap forward in improving the independence and safety of individuals with mobility impairments. These innovative devices not only enhance the quality of life for users by facilitating seamless navigation through complex environments but also offer the potential to revolutionize the way we perceive accessibility and inclusivity in our society. As technology continues to advance, we can expect these devices to become even more sophisticated and integrated into everyday life, further empowering individuals with disabilities and contributing to a more inclusive and equitable world.

References

[1] F. Hill-Briggs, J. G. Dial, D. A. Morere, and A. Joyce, “Neuropsychological assessment of persons with physical disability, visual impairment or blindness, and hearing impairment or deafness,” *Archives of clinical neuropsychology*, vol. 22, no. 3, pp. 389–404, 2007.

[2] A. R. Ulldemolins, V. C. Lansingh, L. G. Valencia, M. J. Carter, and K. A. Eckert, “Social inequalities in blindness and visual impairment: A review of social determinants,” *Indian journal of ophthalmology*, vol. 60, no. 5, p. 368, 2012.

[3] J. E. Harrison, S. Weber, R. Jakob, and C. G. Chute, “Icd-11: An international classification of diseases for the twenty-first century,” *BMC medical informatics and decision making*, vol. 21, no. 6, pp. 1–10, 2021.

[4] S. R. Flaxman, R. R. Bourne, S. Resnikoff, *et al.*, “Global causes of blindness and distance vision impairment 1990–2020: A systematic review and meta-analysis,” *The Lancet Global Health*, vol. 5, no. 12, e1221–e1234, 2017.

[5] WHO, *Blindness and vision impairment*, 2022. [Online]. Available: `https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment`.

[6] D. Pascolini and S. P. Mariotti, “Global estimates of visual impairment: 2010,” *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012.

[7] G. A. Stevens, R. A. White, S. R. Flaxman, *et al.*, “Global prevalence of vision impairment and blindness: Magnitude and temporal trends, 1990–2010,” *Ophthalmology*, vol. 120, no. 12, pp. 2377–2384, 2013.

[8] R. Varma, T. S. Vajaranant, B. Burkemper, *et al.*, “Visual impairment and blindness in adults in the united states: Demographic and geographic variations from 2015 to 2050,” *JAMA ophthalmology*, vol. 134, no. 7, pp. 802–809, 2016.

[9] L. M. Weih, J. B. Hassell, and J. Keeffe, “Assessment of the impact of vision impairment,” *Investigative ophthalmology & visual science*, vol. 43, no. 4, pp. 927–935, 2002.

[10] A. Crudden, “Transportation issues: Perspectives of orientation and mobility providers,” *Journal of Visual Impairment & Blindness*, vol. 109, no. 6, pp. 457–468, 2015. DOI: `10.1177/0145482X1510900604`. eprint: `https://doi.org/10.1177/0145482X1510900604`. [Online]. Available: `https://doi.org/10.1177/0145482X1510900604`.

[11] K. Hara, S. Azenkot, M. Campbell, *et al.*, “Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis,” *ACM Trans. Access. Comput.*, vol. 6, no. 2, 2015, ISSN: 1936-7228. DOI: `10.1145/2717513`. [Online]. Available: `https://doi.org/10.1145/2717513`.

[12] C.-W. Wang, C. L. Chan, A. H. Ho, and Z. Xiong, “Social networks and health-related quality of life among chinese older adults with vision impairment,” *Journal of Aging and Health*, vol. 20, no. 7, pp. 804–823, 2008.

[13] R. L. Brown and A. E. Barrett, "Visual impairment and quality of life among older adults: An examination of explanations for the relationship," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 66, no. 3, pp. 364–373, 2011.

[14] V. Juodžbalienė and K. Muckus, "The influence of the degree of visual impairment on psychomotor reaction and equilibrium maintenance of adolescents," *Medicina*, vol. 42, no. 1, pp. 49–56, 2006.

[15] D. Dakopoulos, "Tyflos: A wearable navigation prototype for blind & visually impaired; design, modelling and experimental results," 2009.

[16] A. Schafer and D. G. Victor, "The future mobility of the world population," *Transportation research part a: policy and practice*, vol. 34, no. 3, pp. 171–205, 2000.

[17] A. D. P. dos Santos, F. O. Medola, M. J. Cinelli, A. R. Garcia Ramirez, and F. E. Sandnes, "Are electronic white canes better than traditional canes? a comparative study with blind and blindfolded participants," *Universal Access in the Information Society*, vol. 20, no. 1, pp. 93–103, 2021.

[18] N. R. Council *et al.*, *Electronic travel aids: New directions for research*. National Academies Press (US), 1986.

[19] A. Delgado-Bonal and J. Martín-Torres, "Human vision is determined based on information theory," *Scientific reports*, vol. 6, no. 1, pp. 1–5, 2016.

[20] J. T. Holladay, "Visual acuity measurements," *Journal of Cataract & Refractive Surgery*, vol. 30, no. 2, pp. 287–290, 2004.

[21] C. Owsley and G. McGwin, "Vision and driving," *Vision Research*, vol. 50, no. 23, pp. 2348–2361, 2010, Vision Research Reviews, ISSN: 0042-6989. DOI: https://doi.org/10.1016/j.visres.2010.05.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0042698910002531.

[22] M. Dougherty, J. Wittenborn, E. Phillips, and B. Swenor, "Published examination-based prevalence of major eye disorders.," 2018.

[23] WHO, *6 who technical consultation on postpartum care - national center for ...* 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK310595/.

[24] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.

[25] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma: A review," *Jama*, vol. 311, no. 18, pp. 1901–1911, 2014.

[26] R. S. Harwerth, L. Carter-Dawson, F. Shen, E. L. Smith, and M. Crawford, "Ganglion cell losses underlying visual field defects from experimental glaucoma," *Investigative ophthalmology & visual science*, vol. 40, no. 10, pp. 2242–2250, 1999.

[27] NORA, *Home*. [Online]. Available: https://noravisionrehab.org/patients-caregivers/facts-and-figures.

[28] K. Kannan, *See the world through the eyes of the blind*. YouTube, 2019. [Online]. Available: https://www.youtube.com/watch?v=PEm3k8UIkbw.

[29] C. Shah, M. Bouzit, M. Youssef, and L. Vasquez, "Evaluation of ru-netra-tactile feedback navigation system for the visually impaired," in *2006 International Workshop on Virtual Rehabilitation*, IEEE, 2006, pp. 72–77.

[30] R. G. Golledge, J. R. Marston, and C. M. Costanzo, “Attitudes of visually impaired persons toward the use of public transportation,” *Journal of Visual Impairment & Blindness*, vol. 91, no. 5, pp. 446–459, 1997.

[31] D. Dakopoulos and N. G. Bourbakis, “Wearable obstacle avoidance electronic travel aids for blind: A survey,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 25–35, 2009.

[32] S. Zafar, M. Asif, M. B. Ahmad, *et al.*, “Assistive devices analysis for visually impaired persons: A review on taxonomy,” *IEEE Access*, vol. 10, pp. 13 354–13 366, 2022.

[33] R. Tapu, B. Mocanu, and T. Zaharia, “Wearable assistive devices for visually impaired: A state of the art survey,” *Pattern Recognition Letters*, vol. 137, pp. 37–52, 2020.

[34] W. Elmannai and K. Elleithy, “Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions,” *Sensors*, vol. 17, no. 3, p. 565, 2017.

[35] J. A. Dowling, A. Maeder, and W. Boles, “Mobility enhancement and assessment for a visual prosthesis,” in *Medical Imaging 2004: Physiology, Function, and Structure from Medical Images*, SPIE, vol. 5369, 2004, pp. 780–791.

[36] W. R. Wiener, R. L. Welsh, and B. B. Blasch, *Foundations of orientation and mobility*. American Foundation for the Blind, 2010, vol. 1.

[37] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, Ieee, vol. 1, 2001, pp. I–I.

[38] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, Ieee, vol. 1, 2005, pp. 886–893.

[39] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[40] R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[41] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.

[42] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks,” *arXiv preprint arXiv:1312.6229*, 2013.

[43] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[44] W. Liu, D. Anguelov, D. Erhan, *et al.*, “Ssd: Single shot multibox detector,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 2016, pp. 21–37.

[45] P. Felzenszwalb, D. McAllester, and D. Ramanan, “A discriminatively trained, multiscale, deformable part model,” in *2008 IEEE conference on computer vision and pattern recognition*, Ieee, 2008, pp. 1–8.

[46] F. S. Bashiri, E. LaRose, J. C. Badger, R. M. D'Souza, Z. Yu, and P. Peissig, "Object detection to assist visually impaired people: A deep neural network adventure," in *International symposium on visual computing*, Springer, 2018, pp. 500–510.

[47] P. Strong, "The history of the white cane," *Paths to Literacy*, 2009.

[48] J. M. Benjamin Jr and N. A. Ali, "An improved laser cane for the blind," in *Quantitative Imagery in the Biomedical Sciences II*, SPIE, vol. 40, 1974, pp. 101–104.

[49] R. Farcy, R. Leroux, A. Jucha, R. Damaschini, C. Grégoire, and A. Zogaghi, "Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view," in *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion*, Citeseer, vol. 12, 2006.

[50] J. Villanueva and R. Farcy, "Optical device indicating a safe free path to blind people," *IEEE transactions on instrumentation and measurement*, vol. 61, no. 1, pp. 170–177, 2011.

[51] B. Hoyle and D. Waters, "Mobility at: The batcane (ultracane)," in *Assistive technology for visually impaired and blind people*, Springer, 2008, pp. 209–229.

[52] B Hoyle and S Dodds, "The ultracane® mobility aid at work training programmes to case studies," *CVHI, Kufstein, Austria*, 2006.

[53] I. Ulrich and J. Borenstein, "The guidecane-applying mobile robot technologies to assist the visually impaired," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 31, no. 2, pp. 131–136, 2001.

[54] T. Ifukube, T. Sasaki, and C. Peng, "A blind mobility aid modeled after echolocation of bats," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 5, pp. 461–465, 1991.

[55] K. Ito, M. Okamoto, J. Akita, *et al.*, "Cyarm: An alternative aid device for blind persons," in *CHI'05 extended abstracts on human factors in computing systems*, 2005, pp. 1483–1488.

[56] J. Hill and J. Black, "The miniguide: A new electronic travel device," *Journal of Visual Impairment & Blindness*, vol. 97, no. 10, pp. 1–6, 2003.

[57] L Kay, "Electronic aids for blind persons: An interdisciplinary subject," *IEE Proceedings A (Physical Science, Measurement and Instrumentation, Management and Education, Reviews)*, vol. 131, no. 7, pp. 559–576, 1984.

[58] J. A. Brabyn, "New developments in mobility and orientation aids for the blind," *IEEE Transactions on Biomedical Engineering*, no. 4, pp. 285–289, 1982.

[59] S. Shoval, J. Borenstein, and Y. Koren, "Mobile robot obstacle avoidance in a computerized travel aid for the blind," in *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, IEEE, 1994, pp. 2023–2028.

[60] S. S. Bhatlawande, J. Mukhopadhyay, and M. Mahadevappa, "Ultrasonic spectacles and waist-belt for visually impaired and blind person," in *2012 National Conference on Communications (NCC)*, IEEE, 2012, pp. 1–4.

[61] S. Bhatlawande, M. Mahadevappa, and J. Mukhopadhyay, "Way-finding electronic bracelet for visually impaired people," in *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, IEEE, 2013, pp. 260–263.

[62] M. R. Strakowski, B. B. Kosmowski, R. Kowalik, and P. Wierzba, “An ultrasonic obstacle detector based on phase beamforming principles,” *IEEE Sensors Journal*, vol. 6, no. 1, pp. 179–186, 2006.

[63] R. Kuc, “Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification,” *IEEE transactions on biomedical engineering*, vol. 49, no. 10, pp. 1173–1180, 2002.

[64] S. Shoval, I. Ulrich, and J. Borenstein, “Navbelt and the guide-cane [o.bstacle-avoidance systems for the blind and visually impaired],” *IEEE robotics & automation magazine*, vol. 10, no. 1, pp. 9–20, 2003.

[65] P. B. Meijer, “An experimental system for auditory image representations,” *IEEE transactions on biomedical engineering*, vol. 39, no. 2, pp. 112–121, 1992.

[66] A. Arnoldussen, “Visual perception for the blind: The brainport vision device,” *Retinal Physician*, vol. 9, no. 1, pp. 32–34, 2012.

[67] N. Molton, S. Se, J. Brady, D. Lee, and P. Probert, “A stereo vision-based aid for the visually impaired,” *Image and vision computing*, vol. 16, no. 4, pp. 251–263, 1998.

[68] R. Tapu, B. Mocanu, A. Bursuc, and T. Zaharia, “A smartphone-based obstacle detection and classification system for assisting visually impaired people,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 444–451.

[69] R. Tapu, B. Mocanu, and T. Zaharia, “Deep-see: Joint object detection, tracking and recognition with application to visually impaired navigational assistance,” *Sensors*, vol. 17, no. 11, p. 2473, 2017.

[70] B. S. Subhashrao, M. Mahadevappa, and J. Mukhopadhyay, *Venucane: An electronic travel aid for visually impaired and blind people*, US Patent App. 14/348,175, 2014.

[71] D. B. Fambro, K. Fitzpatrick, and R. J. Koppa, *Determination of stopping sight distances*. Transportation Research Board, 1997, vol. 400.

[72] M. Rashid, S. Siby, S. PH, *et al.*, “Decreased postural sway in women who are visually impaired: Is it a learned protective mechanism?” *Journal of Visual Impairment & Blindness*, vol. 116, no. 4, pp. 517–525, 2022.

[73] M. A. Goodale and A. D. Milner, “Separate visual pathways for perception and action,” *Trends in neurosciences*, vol. 15, no. 1, pp. 20–25, 1992.

[74] A. D. Milner and M. A. Goodale, “Two visual systems re-viewed,” *Neuropsychologia*, vol. 46, no. 3, pp. 774–785, 2008.

[75] M. Mishkin, L. G. Ungerleider, and K. A. Macko, “Object vision and spatial vision: Two cortical pathways,” *Trends in neurosciences*, vol. 6, pp. 414–417, 1983.

[76] Stereolabs, *Zed 2*, 2019.

[77] I. Krasin, T. Duerig, N. Alldrin, *et al.*, “Openimages: A public dataset for large-scale multi-label and multi-class image classification.,” *Dataset available from https://storage.googleapis.com/openimages/web/index.html*, 2017.

[78] A. Mancini, E. Frontoni, and P. Zingaretti, “Mechatronic system to help visually impaired users during walking and running,” *IEEE transactions on intelligent transportation systems*, vol. 19, no. 2, pp. 649–660, 2018.

[79] B. Mocanu, R. Tapu, and T. Zaharia, "When ultrasonic sensors and computer vision join forces for efficient obstacle detection and recognition," *Sensors*, vol. 16, no. 11, p. 1807, 2016.

[80] R. Jafri, R. L. Campos, S. A. Ali, and H. R. Arabnia, "Visual and infrared sensor data-based obstacle detection for the visually impaired using the google project tango tablet development kit and the unity engine," *IEEE Access*, vol. 6, pp. 443–454, 2017.

[81] A. Burlacu, S. Bostaca, I. Hector, *et al.*, "Obstacle detection in stereo sequences using multiple representations of the disparity map," in *2016 20th International Conference on System Theory, Control and Computing (ICSTCC)*, IEEE, 2016, pp. 854–859.

[82] L. Everding, L. Walger, V. S. Ghaderi, and J. Conradt, "A mobility device for the blind with improved vertical resolution using dynamic vision sensors," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, IEEE, 2016, pp. 1–5.

[83] T. Schwarze, M. Lauer, M. Schwaab, M. Romanovas, S. Bohm, and T. Jurgensohn, "An intuitive mobility aid for visually impaired people based on stereo vision," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 17–25.

[84] K.-F. Lee, "On large-vocabulary speaker-independent continuous speech recognition," *Speech communication*, vol. 7, no. 4, pp. 375–379, 1988.

[85] G. Hinton, L. Deng, D. Yu, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[86] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. L. Y. Bengio, and A. Courville, "Towards end-to-end speech recognition with deep convolutional neural networks," *arXiv preprint arXiv:1701.02720*, 2017.

[87] A. Gulati, J. Qin, C.-C. Chiu, *et al.*, "Conformer: Convolution-augmented transformer for speech recognition," *arXiv preprint arXiv:2005.08100*, 2020.

[88] B. K. Asiedu Asante, C. Broni-Bediako, and H. Imamura, "Exploring multistage gan with self-attention for speech enhancement," *Applied Sciences*, vol. 13, no. 16, p. 9217, 2023.

[89] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[90] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[91] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[92] S. Pascual, A. Bonafonte, and J. Serrà, "Segan: Speech enhancement generative adversarial network," in *INTERSPEECH 2017*, 2017, pp. 3642–3646.

[93] T. Feng, Y. Li, P. Zhang, S. Li, and F. Wang, "Noise classification speech enhancement generative adversarial network," in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, IEEE, vol. 6, 2022, pp. 11–16.

[94] H. Phan, H. Le Nguyen, O. Y. Chén, *et al.*, "Self-attention generative adversarial network for speech enhancement," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 7103–7107.

[95] L. Li, Wudamu, L. Kuerzinger, T. Watzel, and G. Rigoll, "Lightweight end-to-end speech enhancement generative adversarial network using sinc convolutions," *Applied Sciences*, vol. 11, no. 16, p. 7564, 2021.

[96] H. Phan, I. V. McLoughlin, L. Pham, *et al.*, "Improving gans for speech enhancement," *IEEE Signal Processing Letters*, vol. 27, pp. 1700–1704, 2020.

[97] B. McFee, M. McVicar, D. Faronbi, *et al.*, *Librosa/librosa: 0.10.0.post2*, version 0.10.0.post2, Mar. 2023. DOI: 10.5281/zenodo.7746972. [Online]. Available: https://doi.org/10.5281/zenodo.7746972.

[98] O. Buza, G. Toderean, and J. Domokos, "A rule-based approach to build a text-to-speech system for romanian," in *2010 8th International Conference on Communications*, 2010, pp. 83–86. DOI: 10.1109/ICCOMM.2010.5509108.

[99] A. v. d. Oord, S. Dieleman, H. Zen, *et al.*, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[100] Y. Wang, R. Skerry-Ryan, D. Stanton, *et al.*, "Tacotron: Towards end-to-end speech synthesis," *arXiv preprint arXiv:1703.10135*, 2017.

[101] H. Tachibana, K. Uenoyama, and S. Aihara, "Efficiently trainable text-to-speech system based on deep convolutional networks with guided attention," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2018, pp. 4784–4788.