Smart glasses for blind people assistive navigation: an explainable machine learning model

Carlos Romano Gómez

2201887

A thesis submitted for the degree of Master of Science in Intelligent Systems and Robotics

Supervisor: Dr. Sefki Kolozali
School of Computer Science and Electronic Engineering
University of Essex

August 2023

Abstract

In recent times, the demand for effective assistive navigation devices has grown significantly, yet current devices in the industry lack from versatility, explicability, and of an optimal performance in a resourced constrained device. To address this gap, this document introduces the idea of an innovative machine learning model based on tensors to be implemented in an on-board computer. This system would detect objects and give feedback to blind people about their distance, this way, helping them to navigate securely. The proposed system aims to use a Nvidia Jetson Orin Nano computer and a Zed Mini Camera, as these two systems are compatible to run neural networks in a synergic way. Preliminary results of the system present that the model requires much further work as a tensor regression network is a still unexplored path, but present that it can be done by using a classification network and then a sweeper network in the image, but this could produce a big latency in the responses. Regarding the smart glasses, they proved to be comfortable to the users and be able to run a data collection system without any problem. Further work is needed to improve the system and make an interpretable model.

Acknowledgements

I would like to express my deepest gratitude to my family for their unconditional support and encouragement throughout this academic journey. They gave the opportunity to accomplish this academic goal and their belief in my abilities has been a constant source of motivation, and I appreciate their patience and understanding during the long hours of work on this project and stay abroad.

I am also profoundly thankful to my friends for their moral support and kind advices. Their presence has been a reminder that a strong network of relationships is crucial in every aspect of life, including academic pursuits.

I extend my heartfelt appreciation to my supervisor, Sefki Kolozali, for his guidance, insights, and dedication. His expertise has been invaluable in shaping the direction of this project, and his mentorship has been instrumental in fostering my growth as a researcher, I appreciate his support and help deciding my future. His willingness to invest time and effort in providing constructive feedback has been pivotal in refining the project's ideas and methodologies.

I would like to acknowledge the contributions of my lab colleagues, who collaborated with me during various phases of this project. Their insights, brainstorming sessions, and technical assistance have been immensely valuable.

Last but not least, I want to thank all those individuals whose names might not appear here but whose influence and encouragement have contributed to my academic pursuits. Your belief in my potential and your support have been a driving force in reaching this stage.

In conclusion, this project would not have been possible without the collective efforts of my family, friends, supervisor, colleagues, and the wider network of supporters. Your encouragement, advice, and assistance have been indispensable, and I am deeply grateful for your presence in my academic journey.

Contents

List of Figures

CHAPTER 1

Introduction

According to an assessment conducted by the World Health Organization (WHO), an estimated global populace of around 2.2 billion individuals struggles with visual impairments [1]. A considerable amount of this demographic group encounters challenges in navigation, particularly within social indoor settings such as universities, medical facilities, and commercial establishments like supermarkets.

Within these environments, the affected individuals frequently encounter difficulties in acquiring spatial orientation cues and directional information, depleting their ability to determining their whereabouts and optimal routes that will not damage them toward their destinations. This often hampers them and leads to a reluctance to go outdoors, which consequently affects in their social lives and contributing to their isolation [2]. Furthermore, this situation affects people on a large scale, exerting damage on the integral development of young children and compromising the general quality of life experienced by adults. Consequently, along the growth of this high-priority circumstance there has been a noticeable increase in technological innovations. This surge in technological progress has notably generated the conception and construction of novel navigation assistance mechanisms [3] like seen in the recent years. Diverse designs of devices has surfaced with the intent of providing substantive support to visually impaired individuals. Noteworthy among these are devices such as white canes, Global Positioning System (GPS) trackers, smart glasses, and assorted Artificial Intelligence (AI)-driven methodologies [4]. These revolutionary efforts are underpinned by the overall goal of improving the sphere of mobility, fostering a greater sense of autonomy, and fostering a general improvement in the quality of life for people struggling with visual impairments.

Despite their reliability, these devices are not without limitations. For instance, certain methods encounter difficulties in detecting objects, particularly in specific situations, such as head-level objects, a challenge observed in white cane devices [5]. Additionally, it exists a lack of transparency in using deep learning techniques, which poses a significant concern. Users may struggle to comprehend the rationale behind the device's actions, limiting trust and confidence in its use.

Furthermore, developing robust and adaptable algorithms to accommodate various environments requires substantial computational power. Unfortunately, some of the current devices lack the necessary performance and computational power balance, leading to failures when navigating in new environments, restraining them of a better effectiveness and usability.

Addressing these limitations is of much importance to further enhance the functionality and impact of assistive navigation devices for visually impaired individuals. Ongoing research and advancements in AI, Machine Learning (ML), and computer vision offer new opportunities to develop more transparent, efficient, and user-friendly solutions that can empower and provide them with improved independence and mobility to those with visual impairments.

1.1 Objectives

Hence, the prime endeavor of my research endeavors to present a comprehensive solution in the form of a navigation assistive apparatus tailored to surpass these multifaceted challenges. This apparatus will encompass an intricate system, meticulously orchestrated to involve a ML algorithm. This algorithm will be strategically deployed to effectuate object detection within the immediate surroundings, particularly in scenarios relevant to individuals navigating with visual impairments.

To impart a heightened measure of interpretability, the devised system will utilize tensor decomposition, more precisely, the Canonical Polyadic (CP) decomposition approach, along with tensor regression. This choice is propelled by its potential to facilitate an interpretive dialogue between the user and the model, thereby enhancing the transparency of the decision-making process [6]. Notably, this systematic innovation will rigorously pursue equilibrium between computational power and precision, thereby encouraging cautious compensation. The overall goal is to enable timely and astute object detection without compromising the algorithm's versatility in various contextual scenarios.

In the pursuit of operationalizing the envisioned algorithms, the research is willingly to address the ensuing inquiries:

1. Can a interpretable system perform as well as deep learning methods?
2. Can such system efficiently work on a resource constrained device?
3. Can using smart glasses provide a better user experience for blind people?

1.2 Structure

The next sections of this document will explain the following: In chapter 2, a comprehensive review and analysis of relevant literature and existing works related to assistive navigation devices for visually impaired individuals will be presented and discussed. The aim is to identify the strengths and weaknesses in the current devices, providing a foundation for the proposed methodology. In chapter 3 the proposed methodology for the assistive system will be introduced. This encompasses the design and integration of a interpretable algorithm into user-friendly smart glasses, aimed at aiding visually impaired individuals in navigating their surroundings. Also containing the theoretical and technical aspects of the complex algorithm, incorporating Tensor Decomposition techniques. In chapter 4 the outcomes and findings of the implemented assistive system will be presented. Performance metrics, including accuracy, recall, and precision, will be used to evaluate the effectiveness of the algorithm in object detection and decision-making. The discussion will delve into the implications of the results and provide insights into the device's real-world applicability and potential areas of improvement. Finally, chapter 5 will offer a comprehensive conclusion based on the study's outcomes and analysis, summarizing the main contributions of the research and its significance in the field. Additionally, this section will outline potential future research directions and improvements to further enhance the device's capabilities and impact.

CHAPTER 2

Related Work

The trajectory of technological advancement has been directed in a deeper comprehension of the special needs of the visually impaired people. This understanding, in turn, has led to the evolution and creation of superior devices to facilitate their daily experiences. These emerging technologies are meticulously fabricated to enhance the overall quality of life for individuals dealing with visual impairments, extending a diverse array of tools to be used in specific contexts. The main objective is to deliver an efficacious assistance mechanism that accommodates not only the visual limitations but also the spectrum of mobility challenges and unique requisites. These requisites range in all the spectrum, encompassing considerations such as the severity of sensory impairment, the adaptability to outdoor and indoor environments, and the provisioning of feedback via tactile, auditory or other means, to name a few characteristics. This dynamic interplay between technological innovation and personalized user needs underscores a commitment to fostering a more inclusive and enabling environment for the visually impaired community.

The evolution of assistive devices has witnessed remarkable advances over the years, revolutionizing the way visually impaired people navigate their environment. From traditional tools such as white canes to more technologically sophisticated solutions, the quest to improve the independence and mobility of people with visual impairments has stimulated innovation in various fields. These assistive technologies can be broadly classified into two distinct classes [7]: those designed to provide information to users, exemplified by cameras or applications capable of audibly transmitting textual content; and those designed to provide active guidance to the user, predominantly focused on navigation. This document will focus primarily on the latter category, which encompasses a variety of tools, ranging from traditional implements like white canes and service dogs to modern innovations like GPS trackers and smart glasses.

Smart glasses have emerged as a promising avenue, as they incorporate cutting-edge technologies to provide real-time assistance in object recognition, obstacle avoidance, and navigation. Smart glasses, equipped with integrated cameras, have shown their potential to revolutionize the lives of people with visual disabilities. These devices have the ability to identify and provide auditory or visual cues about objects, obstacles, and landmarks in the user's environment, thus offering a higher level of situational awareness and adaptability. These new capabilities not only contribute to safer mobility but also promote a sense of autonomy for the user that leads to greater security and empowerment of the user.

This has been achieved since in recent times, the integration of ML models has ushered in a new era of potential for smart glasses in the field of assistive technology. These models offer a systematic approach to extract meaningful information from complex data, which is particularly advantageous in object recognition and environmental understanding. One such novel model that has been explored has been tensor regression, which unlike conventional algorithms that often struggle with the complexities of real-world data, excels at capturing intricate patterns and relationships within multidimensional data.

2.1 Assistive Navigation Devices for Visually Impaired Individuals

Assistive devices present in different categories, with their specific functions and positioning generating variable responses to certain scenarios, thereby influencing their effectiveness. This categorization defines them into traditional hand-held devices, such as white canes; alternative body-part attachment devices, like chest straps; and head-mounted devices, such as smart glasses.

Within the traditional hand-held category, the most prevalent representative is the white cane followed by the guide dog. The first one, is a device characterized for being economical and lightweight designed for ease of use, primarily necessitating the user's engagement of at least one hand to manipulate it. White canes are proficient at detecting objects with the users help for scanning, particularly those situated below knee level, enabling users to undertake rudimentary object detection within a close proximity [8].

The outstanding popularity of white canes among the visually impaired community arises from their cost-effectiveness, user-friendly design, and widespread accessibility, which sets them apart from other, often more specialized, devices. The main factor contributing to their extensive usage is their availability in diverse configurations and heights, affording individuals the opportunity to adapt their choice to personal comfort and preference.

On the other hand, guide dogs represent a distinct avenue of assistance, also very know and used among blind individuals, as they are characterized by their rigorous training regimens. These highly trained canines serve as astute navigational companions, adept at evading obstacles and aiding individuals in their movement. Beyond their navigation main usage, guide dogs also can be trained to perform additional tasks, but this mostly depends on the dog's training and physique. Some of these activities can be like object retrieval and emotional support. This multifaceted utility underscores the symbiotic partnership that can develop between visually impaired individuals and their guide dogs, presenting a comprehensive solution to various mobility challenges.

Recent advancements have augmented the functionality of these canes by incorporating diverse sensors and supplementary features aimed at enhancing their detection range and accuracy. One of this integration's, the use of ultrasonic sensors presented in [9] and [10] has expanded their capabilities significantly. These ultrasonic sensors enable the detection of objects up to a meter's distance from the user and extend their effectiveness to waist-level object detection. They also present a sound feedback capable of sending the necessary information to the user in a timely manner about the detected objects, leaving a considerable window of time for the user to respond.

Nonetheless, these traditional aids are not without their limitations. They present several drawbacks, but the most notable one is based on their efficacy in detecting obstacles within the immediate vicinity and certain height. As these aids typically do not present a broad spatial field for object detection, they are limited to detect objects up to certain height that normally do not surpass the waist. In this way, they overlook potential impediments positioned at head level or originating from other parts but the front [5]. This restricted scope can leave blind individuals vulnerable to obstructions that might be dangerous in a daily basis.

Furthermore, a significant constraint linked to these aids is the inverse relationship between their detection range and the user's walking speed since the reliance on distance for detection can inadvertently reduce the user's pace during navigation. This phenomenon surges from the confidence of the user in the device, as they tend to maintain a slower gait in a device they do not trust to ensure timely detection and avoidance of objects. The reduced walking speed can potentially hinder the user's freedom of movement and efficiency during daily mobility tasks, thereby underscoring a pivotal challenge associated with these traditional assistive devices.

In spite of the many advantages associated with utilizing trained guide dogs, a significant drawback revolves around the expensive financial investment required for their complete training. The costs involved in training these animals often reach levels that exceeds the income for many individuals, meaning that they are not for everyone. Moreover, the extensive training period, spanning months, further compounds the challenge, as it precludes a swift and accessible solution to the needs of visually impaired individuals seeking immediate assistance for their navigation challenges. The combination of formidable costs and extensive training duration does not make this approach suitable for providing effective and efficient assistance to this specific community without incurring such substantial resource and time commitments.

The subsequent classification encompasses the domain of body-attached devices. This category embraces an array of intricately engineered devices that harmonize software and hardware components to yield instruments capable of detecting objects across diverse spatial configurations and forms. Typically, these devices amalgamate intricate sensors and sophisticated software algorithms to facilitate object detection. This confluence of technologies yields a diverse array of devices, tailored to specific scenarios and individual needs.

Examples of such devices span a wide spectrum, encompassing Haptic feedback [11, 12], Global Positioning Systems (GPS) [13, 14, 15], water sensors [16], cameras, and even assistive apparatuses designed for oral use [7] are commonly employed elements within this category. All of these devices offer unique characteristics that help visual impaired people in one way or another.

Exploring deeper into each of these examples, let's begin with haptic feedback devices. These devices operate by utilizing ultrasounds or other similar components to detect objects in the environment. They then transfer the information to the user through the sense of touch, often employing distinctive vibration patterns like Morse code or specific tactile patterns to send information about the surroundings. A notable advantage of this approach is that the user do not lose information that comes from other sources, like it can be audio cues [11]. However, these devices come with significant drawbacks. One major challenge lies in the user's need to memorize the specific tactile patterns, introducing an element of reliance on memory, and a risk in confusing the user and lead them to danger. Additionally, the effectiveness of the device is contingent upon the rate at which it can transmit Morse code or other patterns, as the user must interpret these signals swiftly to avoid potential accidents.

Moving on, we encounter GPS systems and water sensors, renowned for their widespread adoption, cost-effectiveness, and portability. The first one, offers a system with the capacity to track an individual's location across most parts of the world, thereby facilitating the monitoring of the user's navigational trajectory for future reference. And the second one is primarily a sensor that can detect if there's any liquid spill in the path the user walks, which can avoid any slip accident. However, even though it is notable that GPS systems primarily excel in location tracking and path tracing, and water sensors can detect liquids with ease, they inherently lack the ability to detect obstacles, detect the facing direction of the user or proactively prevent accidents due to their sole focus on one task, geographic positioning [15]

for GPS and water detection for water sensors. Consequently, these sensors are frequently integrated as supplementary features within devices, rendering them less pertinent in the context of object detection.
Regarding the general limitations that this type of devices present, a predominant one, inherent to these devices, is their different alignment with the everyday experiences of visually impaired individuals, as these devices design and testing is often concentrated on carefully designed paths tailored for experimental purposes, which may not accurately replicate the diverse challenges and scenarios encountered by blind people in their daily lives. Consequently, the controlled environments used for testing can inadvertently downplay the devices' real-world applicability and practicality, potentially leading to an overestimation of their effectiveness.
Finally, another significant drawback, that is also shared by handheld devices, relates to their limited spatial range, making them ineffective in detecting objects situated above the device's height. This deficiency in perceiving obstructions at head level poses a serious safety concern, potentially subjecting users to unforeseen hazards. The incapacity to adequately identify obstacles positioned at head height not only endangers user well-being but also casts doubts on the overall effectiveness of these devices. Unfortunately, accidents involving inadvertent head collisions are prevalent, underscoring the critical need for comprehensive design considerations that prioritize both user safety and device utility.
The final category encompasses the head mounted devices. This classification predominantly works around the use of cameras, both monocular and stereo, coupled with computing components such as computers or micro controllers to employ image processing techniques for effective object detection and interpretation.
Within this context, the focal point turns to smart glasses, which constitute a prominent subset of head-mounted assistive devices. These glasses have gained notable attention due to their remarkable potential to augment navigation and interaction functionalities for individuals with visual impairments. Distinguished by their adaptability, these devices can be secured through a glasses frame or adjustable straps, pairing to the unique requirements and preferences of each user. Also, this technique has demonstrated considerable advancements in recent years, proving versatile across an array of tasks including obstacle avoidance [17, 18, 19], text-to-speech capabilities [20, 21], and even currency detection on banknotes [22, 23]. Another notable attribute of these smart glasses lies in their capacity to empower users with uninhibited mobility. Similar to body-mounted devices and in contrast to handheld aids, these glasses unburden the user's hands, facilitating a wider spectrum of tasks to be performed with ease. This facet significantly enhances the user's engagement with their surroundings, fostering an increased sense of independence and seamless navigation.
However, within the realm of smart glasses, a big variety of devices exist, each potentially superior to others based on their functionalities or hardware configurations. For instance, commercially available smart glasses, despite boasting an array of features, often prove unsuitable for blind users due to their lack of accessibility-oriented attributes. Moreover, their exorbitant cost can render them inaccessible to a broad audience. On the other hand, many smart glasses prototypes in the industry utilize monocular cameras for object detection. While functional, this approach comes with its own set of limitations. Monocular cameras tend to yield imprecise distance measurements, or they may consume greater computational resources due to the necessity of more intricate calculations compared to setups using stereo cameras or RGB-Depth camera configurations [17]. The merit of employing such camera types lies in the fact that, with knowledge of the focal length and the separation between the two lenses, distance calculations can be conveniently derived using the formula $D = (WxF)/P$, where D signifies the distance, W denotes a

known width, and P stands for the width in pixels. Then, the most important thing is, as they use cameras to detect objects, they are based on computer vision algorithms to locate the objects in the image and give feedback depending on what they locate. So the next section will delve more profundly in this types of models.

2.2 Machine Learning in Assistive Navigation

Assistive devices use ML models to alert users to potential hazards, particularly those equipped with integrated cameras that rely on computer vision techniques. While ML models might exhibit initial low accuracy, their potential for significant enhancement is undeniable. Several strategies can be employed to bolster their performance. Transfer learning, for instance, involves using a pre-trained model as the foundation and transferring its features and weights to a new model, expediting training time and boosting accuracy. Collecting more diverse training-testing data is another avenue, though it requires caution to avert overfitting. Moreover, introducing dropout layers, which temporarily deactivate certain neurons, can mitigate model complexity and overfitting risks.

The efficacy of these models hinges on hardware choices, like mono or stereo cameras, as well as the quality of their initial training data. Often, researchers opt for a pretrained model as a starting point, subsequently fine-tuning it using transfer learning on their specific datasets. A significant benefit of widely used ML algorithms, such as CNN, is their capacity to accommodate an expanded array of detectable classes, rendering them versatile across various scenarios.

Taking a closer look at specific studies like those found in references [24], [25], and [26], they propose systems integrating smart glasses furnished with stereo cameras. While their models are based on pretrained counterparts and can identify numerous categories, achieving an accuracy of approximately 70% it may not meet the ideal requirements posed by visually impaired users. These studies underscore the ongoing quest for more refined and accurate ML models in the realm of assistive devices for the blind.

2.3 Tensor Decomposition in Machine Learning

In order to came up with an innovative algorithm, it is necessary to comprehend the alternatives that exists to the traditional neural networks. To this end, an exhaustive exploration of scholarly literature in the domain of neural networks has substantively informed the conception of this labor. Before continuing, an integral part of this discourse is the definition of the term "tensor", which serves as a fundamental unit of information encompassing complex data constructs. Tensors make it easy to analyze multidimensional data, thus unraveling latent patterns that can evade detection when analyzed within a solitary dimension. For example, this concept finds prominent application in the realm of multi modal data analysis, where tensors, often recognized as matrices, skillfully navigate various dimensions to extract information. Illustratively, in the context of images, tensors effectively encapsulate the trifecta of height, width, and color dimensions, facilitating an understanding of the data set and what it contains. Applying this knowledge into a neural network environment, it surges a Tensor Regression Network (TRN), a recent addition to neural network architectures, that shares a construction resemblance with Convolutional Neural Networks (CNN). Like CNNs, TRNs capitalize on the input of three-dimensional tensors within their convolutional layers. This framework retains a multilinear structure. However, a distinctive departure from CNNs arises, as TRNs do not interface with a fully connected layer; instead, they are linked to a tensor contraction layer. This configuration serves to uphold the essential attributes of multilinearity and multimodality innate to the system, simultaneously decreasing the number of requisite parameters. This class of networks, although new in the literature, has begun to capture attention. Research, such as the work by [27], asserts that these networks can be harnessed for large image classification. The strategy involves leveraging a pretrained model and substituting the final layer with a Tensor Regression Layer. This adaptive approach culminates in improved accuracy and a more compact space requirement, often yielding up to a 25% reduction in space consumption. This tensor regression layer its just a tensor building and decomposition to decrease the dimensionality but made into a layer that the network can read.

This tensor decomposition concept is a relatively new in the area of neural networks, leading the idea to being applied into neural networks to obtain different results, that can go from making faster networks to making different feature extraction algorithms. In the realm of tensor decomposition encompasses two primary methodologies: Tucker Decomposition and Canonical Polyadic (CP) Decomposition, often referred to as the higher-order Singular Value Decomposition (SVD) and the expression of a tensor as a summation of finite rank-one tensors, respectively [28]. While these techniques share certain similarities, they also bear distinctive attributes. For instance, Tucker decomposition permits the accommodation of varying quantities of factors across tensors, a departure from the more rigid CP decomposition [29]. It is noteworthy, however, that Tucker decomposition can encounter diminished efficiency as the order of tensors escalates [30]. It is imperative to understand that the integration of tensor regression and object detection within a neural network constitutes a pioneering exploration in the field. Thus far, no much information of such a fusion has emerged, leaving the future researches in a completely new environment. But it is know that in order to execute this fusion, a crucial requirement is the generation of a tensor encompassing essential information as object coordinates within the image alongside their corresponding classes and feed it to the modified TRN so that it can make the regression of new boxes along the expected classes. Alternatively, a different route requires crafting a robust tensor network that initially undertakes image classification, followed by a subsequent network that iteratively processes the image to locate the object boxes and subsequently approximate the associated class. Nonetheless, this latter approach could potentially induce a significant latency, making it unreliable for real-time application. Given the context of aiding visually impaired users, such latency could yield unfortunate consequences.

CHAPTER 3

Methodology

This document presents a comprehensive proposal describing a revolutionary smart glasses system designed to boost the navigation capabilities of individuals afflicted by visual impairments. The core of this proposal resides in the incorporation of an innovative algorithm synergistically built upon a 3D printed frame, a depth camera and an onboard computational unit. The primary ambition is to generate a sophisticated apparatus equipped with the ability to discern objects within the user's proximate vicinity, subsequently alerting the user concerning the identified object as well as its spatial proximity. The schematic delineating the operational intricacies of this system is depicted in Figure 3.1 and will be explained later in each subsystem's chapters.

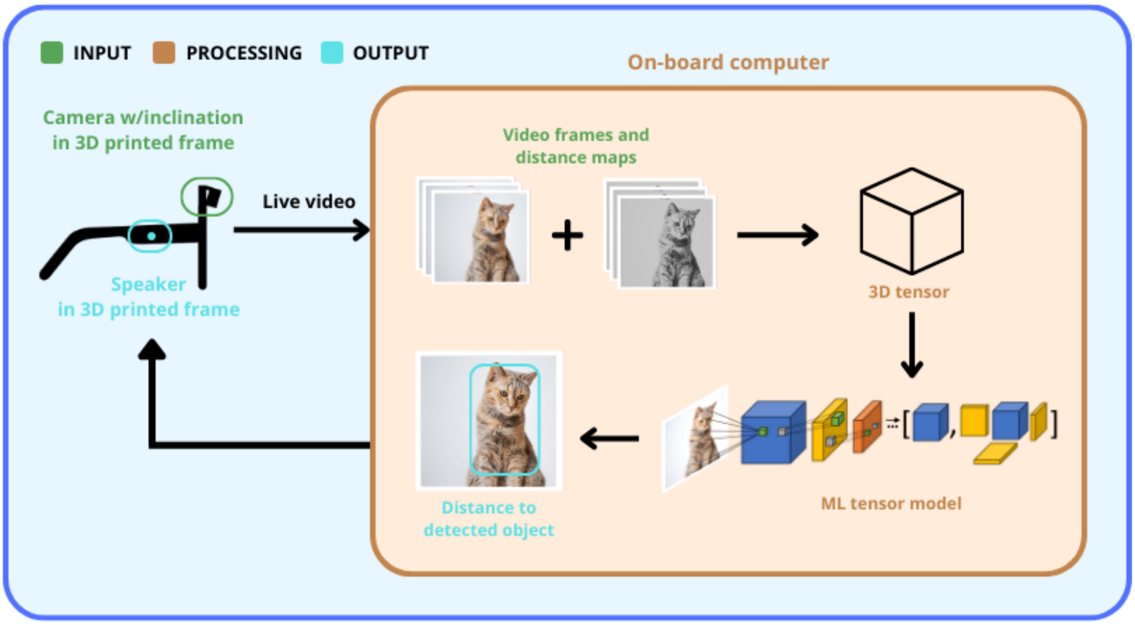

Figure 3.1: System architecture. An image its capured by the camera, passed as frames to the ML tensor model and onced processed, feedback is sent to the user.

In essence, the envisioned system encompasses two vital phases that function in unison to grant its operational efficacy. The software component, constituting the ML algorithm, stands by an explicable algorithm that ingeniously strikes a harmonious equilibrium between computational efficiency and power.

This algorithm assumes a crucial role in orchestrating multifarious tasks spanning from data collection and model training to real-time user feedback, which are the system's core functionalities. This facet prioritizes a proper balance between computational robustness, precision, and system performance. The hardware facet, on the other hand, revolves around the fabrication of the glasses frame tailored to accommodate visually impaired users, alongside the provisioning of a portable element for the onboard computational entity. This facet, resonating with user-centric considerations, underscores the importance of user comfort and operational convenience in the overall design.

3.1 Smart Glasses

In order to provide the proposed system with the necessary portability and the perfect integration for users, it is necessary to adopt an innovative strategy. Central to this strategy is the integration of a meticulously designed 3D printed glasses frame, which assumes the role of containing the camera apparatus for detecting objects within the wearer's immediate view. Importantly, the inherent versatility of this design is highlighted by the provision of strategically allocated spaces, prepared to accommodate potential sensor integration's, ensuring a forward-thinking approach to technology augmentation.
Of great significance is the configuration of the camera's mounting. Positioned at the forefront of the frame, this apparatus assumes an inclination of a predetermined angle. This particular configuration, denoted by the angle θ, serves as a critical determinant in the camera's viewing spectrum. By properly tilting the camera, the resulting panoramic visual expanse encompasses the frontal domain, amplifying the detection range manifold. This efficacious design empowers the camera to discern objects occupying varying altitudes, including those at head level and ground level. The illustrative elucidation of this design elucidation is graphically depicted in the Figure 3.2.

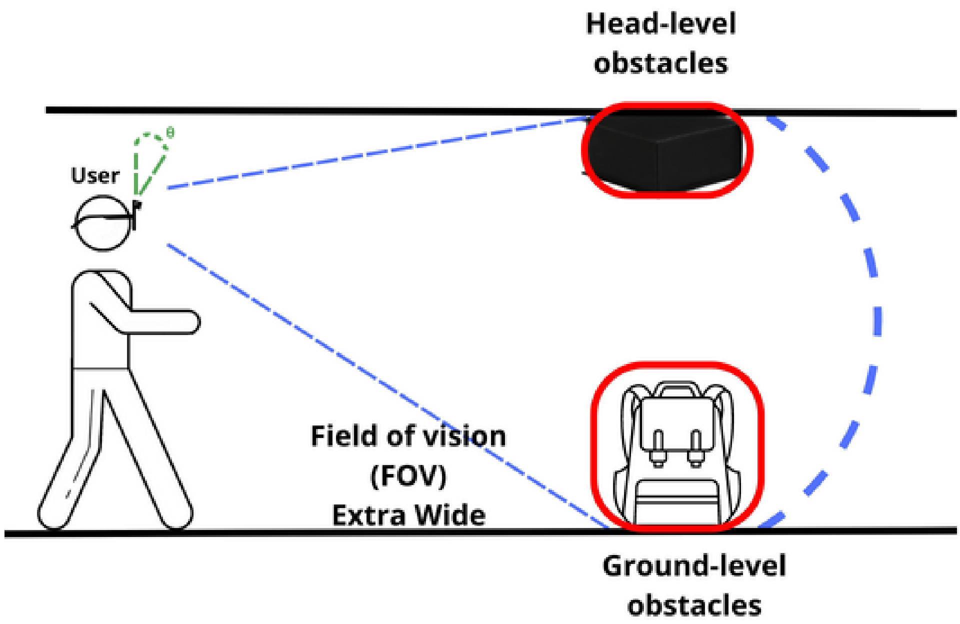

Figure 3.2: A side view of the system. The system will be able to detect objects in a head level and ground level, due to an inclination of the camera in the frame.

In this project, the ZED Mini camera, a product of the Stereo Labs company, was employed. This camera system encompasses two distinct cameras, each characterized by unique functionalities. The

first camera claims a resolution of up to 2.2K and exhibits a variable frame rate of up to 100 Frames Per Second (FPS) depending on the chosen resolution. The second camera, in addition to its imaging capabilities, possesses the capacity to perform depth measurements at a range of up to 15 meters [31]. The careful selection of this device arise from its multifaceted attributes, including its compact form factor.
The intrinsic versatility of the ZED Mini camera is a defining factor in its selection. This versatility is exemplified by its efficacy in diverse data acquisition scenarios. The camera accommodates data acquisition via recording, facilitating the accumulation of training data, as well as real-time data acquisition via streaming, thereby catering to continuous user engagement. Moreover, the camera's inherent adaptability extends to the domain of image quality. The ability to manipulate image quality, whilst influencing the rapidity of user response, stand out the dynamic responsiveness to this system.
It should be noted that the camera is of greater importance since it also has auxiliary capabilities. Among these features are motion sensors, that is, sensors such as gyroscope and accelerometer to detect the movement of the camera, and also has pre-installed object detection models. While these ancillary features remain secondary to the current project goals, their potential addition holds promise for subsequent iterations of this technological innovation.
Similarly, to achieve the required computational prowess while preserving the portability of the device, the adoption of an on-board computer or carrier board was considered appropriate. Here, an onboard computer refers to a processing power unit embedded within a carrier circuit board along with all the essential laptop-like components, including RAM, cooling mechanisms, and add-on components. Notably absent are peripheral devices like monitors, mice, and keyboards. But in return, a carrier board can encompass ports for USB or Ethernet connectivity and come with an operating system pre-installed, facilitating optimal functionality.
In light of this, the Jetson Orin Nano on-board computer by Nvidia surged as the chosen solution. This compact computing device, ideal for the creation of entry-level AI-powered robotics, intelligent drones, and smart cameras [32] akin to the one envisaged within this document, was deemed optimal. The selection of the Jetson Orin Nano was primarily driven by Nvidia's offerings in terms of video components, capable of expeditiously processing multimedia data. Furthermore, this diminutive computing marvel aligns seamlessly with the implementation of AI algorithms, offering an interface for swift and efficient integration.
Given these considerations, in the midst of the large number of specifications offered by both devices, this project has opted for the adoption of LOSSLESS image compression for the camera, specifically at a scale of 2:1. This choice emanates from the innate properties of the specific Nvidia Jetson model used in this context, due to this particular model lacks a hardware accelerator, which prevents the adoption of more aggressive image compression techniques that can yield higher compression rates. Consequently, a setting of 60 frames per second (FPS) along with a high-definition (HD) resolution of 720 was identified as the fastest configuration, although the resulting operating speed measures approximately 10 FPS. This noticeably leisurely pace, however, it also means that the model will have the ability to encapsulate substantial information within each frame, a critical attribute that enables efficient user alerts.
Although the above deliberations, fundamentally address the viability of implementing an algorithm within resource-constrained device, to address the user comfort, the merits of harnessing 3D printing technology have been studied. This strategic choice arises from multifaceted considerations. Notably, the economic advantages inherent in 3D printing, coupled with its inherent adaptability to personalized customization, render it an optimal avenue for both prototyping and potentially for the eventual definitive

model. Furthermore, the innate modular nature of 3D printing facilitates easy incorporation of modular components, thereby harboring latent potential for future enhancements. This portends the prospect of seamlessly integrating adjunct accessories, including microphones, to enhance the object detection capabilities, as previously elucidated in this exposition. Additionally, the incorporation of 3D printing technology has led to the creation of a protective casing, cleverly designed to house the on-board computer and accompanying power supply. This strategic innovation is of great importance as it fulfills the dual purpose of improving security and guaranteeing the longevity of the system, increasing its durability and resistance.

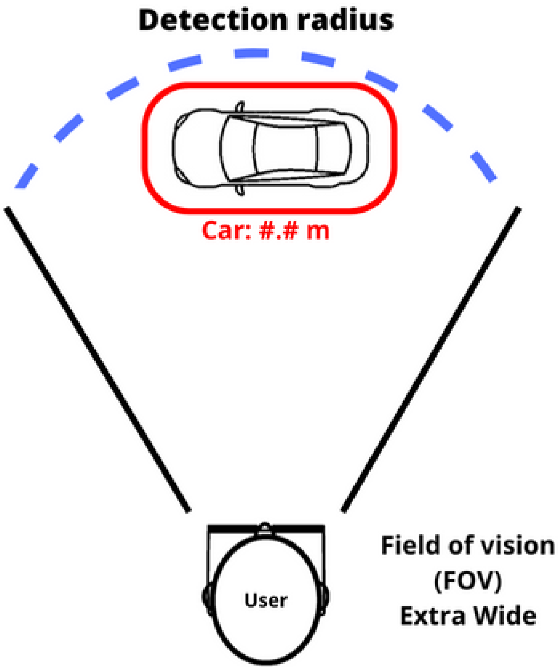

Figure 3.3: A top view of the system. The system will be able to detect objects in front of the user in a wide view angle and give feedback about the distance of the detected objects.

3.2 Machine Learning Algorithm

The fundamental idea underlying the formulation of this algorithm resides in the pursuit of innovative methodologies to develop a novel problem-solving prototype, characterized by an inherent capacity for a custom-made explanation to user comprehension. Particularly, the conception that emerged is stated upon the combination of factor decomposition and a neural network grounded in tensor regression. Notably, this union is presented as a novel approach, prepared to offer a greater degree of versatility in the assimilation of information within neural networks. This greater versatility is underscored by the inherent structural variability of tensors, spanning a wide spectrum of dimensional manifestations.
As highlighted above, the training phase of the algorithm will be done offline, using recorded videos obtained from ZED's camera. However, during the testing phase, a real-time streaming approach will be adopted. This strategic difference in data acquisition methodologies facilitated the creation of distinct codes capable of discerning various attributes, such as color images, spatial depth representation, and individual depth values. This way of obtaining precision facilitated the analysis of simultaneous data, leaving aside concerns related to the sparcity within the data as well as a easy way to form tensors with

the information of the image and distance.
This creation ends in the formulation of a height tensor, denoted as H, which naturally corresponds to the vertical dimension of the image. Within the scope of this experiment, where a resolution of 720 pixels was maintained, the height tensor, H, is inherently held at 720 pixels. Whereas, the width tensor, W, will have the 1280 pixels in relation to the width of the image. In this could also be done a 2 times W width tensor, that emerges as a composite construction, divided to encompass two discrete domains. The first part of W would corresponds to the values that encapsulate the color image, while the second facet would correlate with the values that delimit the representation of spatial depth. In the end, this composite tensor will hold a total size of batchx720x1280x3, that correspond to the color image size and number of images, encapsulating the information inherent in the algorithm's data construct. This structural framework finds a visual appearance in Figure 3.4.

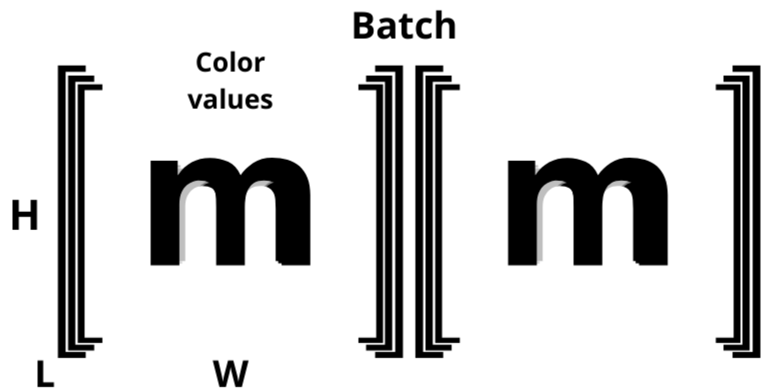

Figure 3.4: A visual representation of the input tensor.

Operational within the algorithmic framework, the proposed system will try to implement the architecture constituted by a tensor regression neural network. Its preliminary structure involves convolutional layers, designed to extract salient features related to each image, and unlike conventional processes, the back layer abandons the standard flattened configuration. Instead, it adopts a new tensor contraction layer, naturally calibrated to preserve multilinear attributes related to the tensor data. This unique architectural choice encourages the preservation of the multilinearity of the tensor and avoiding the possible sparcity and loss of information. This preservation can also be effected through tensor decomposition, thus effectively smoothing the number of tensor ranks, an approach supported by its ability to improve computational efficiency while preserving essential nuances in the data. Towards the end of the traversed path of the network, the result is manifested as a product that involves the resulting tensor, synthesized through the algorithmic process, together with a low rank tensor, which contains the corresponding weighting coefficients [33]. This product, similar to what happens in a conventional neural network, is subsequently subjected to the backpropagation process, a major component for solving the optimization of neural networks. This procedure, which is run iteratively over several epochs, leads to refinement and alignment of the network parameters, ultimately culminating in the achievement of optimal performance. It is imperative to highlight that the conceptualization of this novel neural network is grounded in the current progress within the realm of tensor networks. Nevertheless, it's essential to acknowledge the possibility that practical implementation might face certain constraints. Should such limitations arise, an alternative approach involving a more foundational neural network could be considered. This alternative might leverage conventional CNNs, while meticulously preserving high precision. This strategic direction aims

to ensure the viability of the smart glasses for visually impaired individuals.
To acquire the necessary data to fuel the model, a streamlined approach involving the integration of the lenses and the Jetson computer has been devised. The process begins by transitioning the Jetson computer into Headless mode, effectively eliminating the need for a monitor. This facilitates remote control via an external computer, in this case, a laptop. To power the Jetson, an external battery is employed. The entire setup is then compactly assembled, affixed to the lenses, and securely placed within a backpack, ensuring portability and convenience for field usage.
Upon powering up the Jetson, it will become visible on the laptop's serial ports. This initiates the option to establish a connection through a third-party application called Putty (Figure 3.5). Through this tool, a USB linkage is established between the two computers, enabling the Jetson's control via a terminal interface.

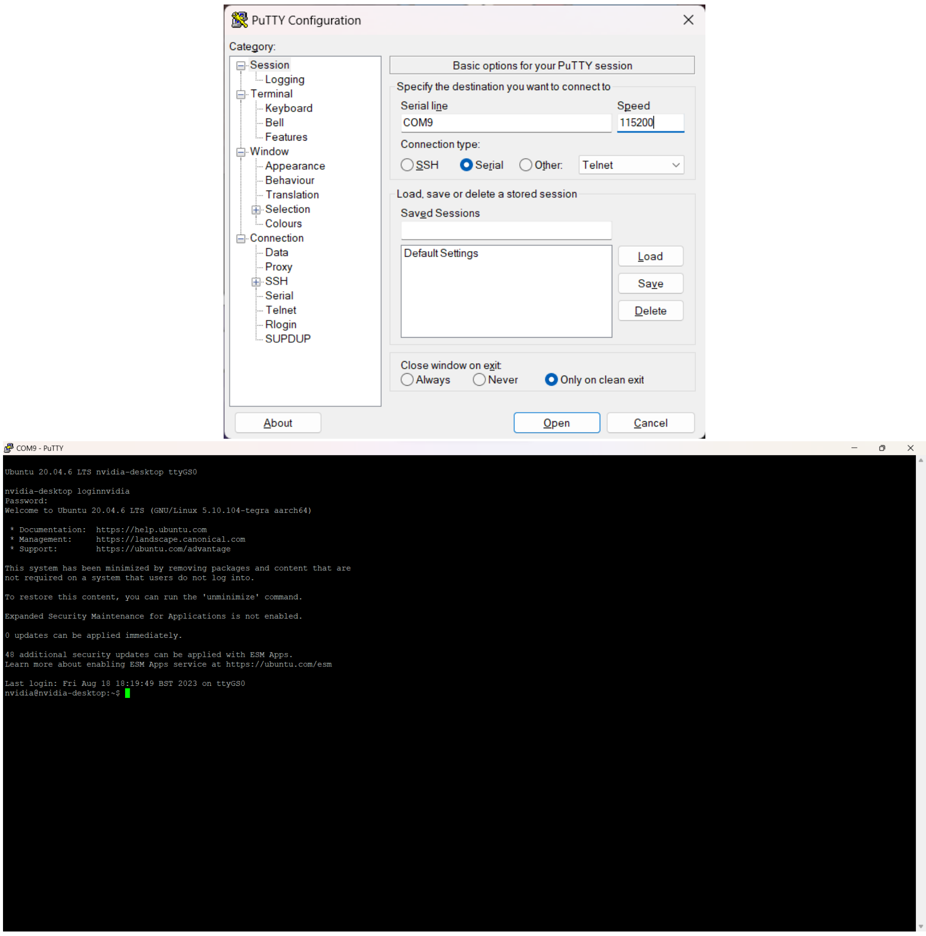

Figure 3.5: A view on how the computer can be controlled from the laptop (Above - Putty app, Below - Terminal from laptop).

With the connection established, the subsequent step involves executing the data collection process. This entails traversing the meticulously chosen paths previously designated for recording. These paths are thoughtfully selected to maximize the capture of intricate details and objects. The selection and representation of these paths can be observed in the illustrative Figure 3.6. This strategic approach aims to acquire a comprehensive dataset that encompasses diverse scenarios, essential for robust model training and performance evaluation.

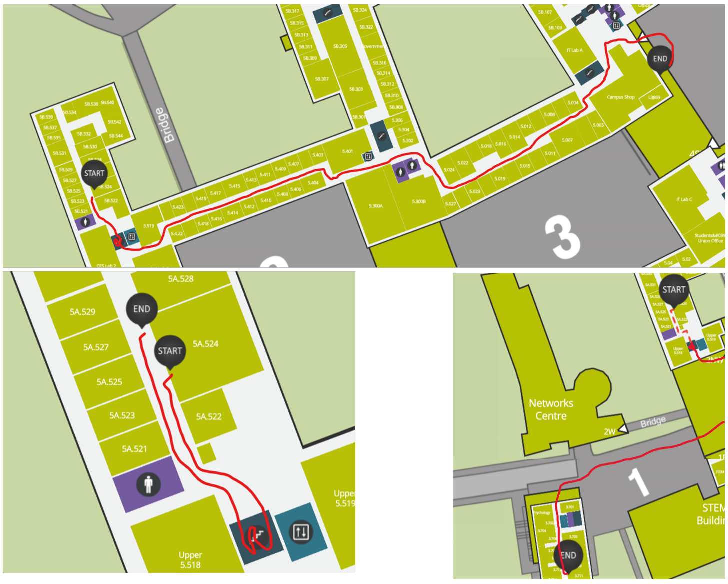

Figure 3.6: An approximate aerial view of the paths chosen, each starting on the same point.

3.3 Project and Risk Management

Throughout the course of this project's development, a range of tools was employed to ensure consistent progress and facilitate problem-solving. This section dives upon these methodologies and offers an introduction and elucidation of the risk management strategies involved in conducting the test runs. These measures were taken to safeguard the well-being of participants and all personnel involved in the testing while concurrently upholding the reliability of the gathered data, pivotal for subsequent analysis. Firstly, the tools and resources used were given by the University of Essex as part of the course structure. Predominantly, this encompassed licenses and software utilities such as GitLab and Jira. Both of these platforms are engineered for effective project management, offering collaborative access to the project stakeholders. GitLab, the first of the two, serves as a collaborative coding environment where in users can establish repositories, either public or private, for code storage and version control. In this context, it played a primordial role in maintaining a constant line of communication with the project supervisor, facilitating the exchange of feedback and updates. However, due to repository capacity constraints imposed by the university, a transition to GitHub was necessary. The entire repository was a personal repository inside this platform similar in functionality but more tailored towards individual code repository management as opposed to collaborative coding environments. The repository link housing the final codes and documents was shared with the supervisor and is accessible at: https://github.com/A01632681Carlos_Romano_Gomez_MSc_Dissertation.

Similarly, the second tool, namely Jira, assumed a central role in documenting project advancements and task distribution. Its primary utility was to provide an extensive record of project progression, meticulously detailing the ongoing tasks and their respective statuses. This facilitated effective communication with the project supervisor, providing them an understanding of the predominant challenges and focal areas at any given moment. Notably, Jira played an instrumental role in discerning both the evolution of the project and any potential opportunities for optimizing time management.

Next, in order to comprehend the testing procedures employed to the project it is necessary to understand the risks associated with the project. Notably, the testing phase entailed a non-blind user employing the smart glasses model to record data across the university campus. Given the imperative of portability, meticulous precaution was essential concerning the on-board computer and its external power source. These components, being provided by the University, underwent rigorous evaluation by the IT department to ascertain their electrical safety, a scrutiny which they successfully passed. However, the external battery exhibited a tendency to overheat, thus presenting a potential hazard during transport. After consultation with the project supervisor, a prudent decision was reached - placing the battery within an empty backpack beneath the computer. This strategic placement aimed to mitigate risks, enabling swift removal of the backpack in case of a fire, thereby minimizing potential damage to the computer. Likewise, the computer was encased within a "cage." This encasement not only limited the exposure of various computer elements but also upgraded the system's resistance against inadvertent impacts, simultaneously ensuring both portability and security.

Now, during the testing phase, achieving an environment similar to that experienced by visually impaired individuals was imperative. While options such as closing one's eyes or using a blindfold were conceivable, these possessed potential safety hazards. Consequently, a more judicious approach was adopted, simulating blindness while maintaining an accompanying individual to record data and provide assistance or guidance in the face of potential dangers.

Furthermore, it's important to address the ethical considerations tied to recording individuals within the university campus. Recognizing the possibility of inadvertently recording students or personnel during operational hours, precautions were taken to maintain privacy and prevent any concerns among those recorded. To this end, a distinctive sign displaying contact information was worn during test runs, facilitating contact and resolution of any potential queries or complaints regarding recorded imagery. It's imperative to underscore that both the recording hours and paths were meticulously selected, with extreme cautious to collect valuable data while safeguarding the privacy and image of individuals.

CHAPTER 4

Results and Discussion

The subsequent section presents a comprehensive overview of the initial results achieved during the course of this study. The undertaken efforts are outlined, shedding light on the attained outcomes and insights gained. Although the main objective was not achieved, the other two objectives were completed and it is hoped that the documented findings contribute to a foundational understanding of the challenges, achievements, and potential pathways for future advancements in the field. It is important to acknowledge that the incompleteness of the project may have implications on the comprehensiveness of the results; however, the insights gained thus far provide valuable directions for further investigation and refinement. With these considerations in mind, the initial outcomes refers to the system's portability. As previously mentioned, the smart glasses were designed using a 3D modeling approach, followed by 3D printing, a strategy that provided considerable configurability in alignment with user-specific requisites, particularly that of the participant undertaking the recording trials. In example of this configurability, the first prototype presented less-than-ideal outcomes due to its oversized frame (Figure 4.1), but this informed subsequent enhancements, resulting in an improved fit for an average user's facial contours. This design evolution further facilitated provisions such as the incorporation of hooks, strategically positioned on the sides, to accommodate supplementary devices or facilitate cable organization, notably from the camera to the computer. A visual representation of the system is featured below in Figure 4.2.

As seen in the Figure 4.2, the camera had the inclination to aptly perceive objects situated on the ground, but also at head level, thanks to the wide-angle configuration facilitated by the VGA resolution. Also, to ensure camera's safety and stability, it was secured by a strap to the frame guaranteeing uninterrupted functionality without causing any discomfort to the user. To counteract the forward-pulling weight of the camera, the glasses had a strap to secure the glasses to the head, avoiding them to fall. With this strap in place, which could not be adjusted tightly, the glasses did not move or fall. Despite the weight of the camera, the user reported minimal inconvenience, suggesting that the average system weight was well-tolerated. The images also feature the specially designed cage intended to securely hold the computer. This precisely crafted enclosure was made to perfectly accommodate the computer, preventing any internal collisions or the possibility of dislodgment and fall. It is worth mentioning that some of the recording tests were made in conjunction with an audio analysis project, exploring the possibility of future joint work. These tests were almost successful since it was possible to collect data in contiguous times, but not at the same time, so if it were to be developed, more future work would be needed. In the same way, the

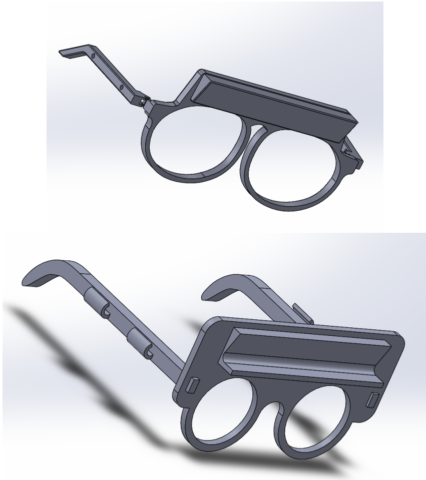

Figure 4.1: A comparison between glasses models (Old - above, New - Below).

Figure 4.2: A visual representation of the system built.

codes used to obtain the joint information can be found on the project's GitHub.

Conclusively, Figure 4.3, illustrates that the system demonstrated the capability for remote operation through the utilization of an alternate computer to execute the necessary codes. Remarkably, this setup eliminated the requirement for a monitor, mouse, or keyboard, thereby enhancing its portability. The onboard computer could be powered by its battery while simultaneously connected to the laptop, enabling seamless control and operation of the entire system. This innovative approach not only ensured portability but also highlighted the adaptability of the system to various operational scenarios.

However, in contrast to the successful implementation of the portable system, the machine learning model rooted in tensor-based methodologies proved to be a challenge warranting further investigation. This challenge stems from the absence of comprehensive resources in the field of networks. The initial attempt involved the integration of a machine learning model featuring tensor regression layers. However, a critical limitation emerged as there is a lack of well-established libraries capable of seamlessly connecting tensor regression layers with existing neural network models. Notably, recent research, exemplified by [27], introduces a class designed to incorporate a Tensor Regression Layer into a Convolutional Neural Network (CNN) model developed using TensorFlow. This particular model, premised on a pretrained classification architecture, remains incompatible with an object detection framework. The inherent challenge lies in

Figure 4.3: A visual representation of the portable size system.

transforming the tensors to a higher-order structure capable of accommodating object coordinates and corresponding classes, a formidable task that necessitates comprehensive exploration and innovation.
Considering these challenges, to ensure the realization of a comprehensive assistive system tailored for visually impaired individuals, an alternative approach is proposed. Given the proven reliability of CNN models, the proposal revolves around the integration of a pretrained CNN model. While the initial aspiration of incorporating tensor regression models might have encountered technical barriers, the new proposal emphasizes a seamless synergy between the two hardware components, ensuring efficient and timely object detection to provide optimal guidance for users. To construct this alternative solution, two feasible options emerge.
The first require training a CNN model from scratch using the collected and labeled data. However, this avenue is deprived by the potential duration of training, which could take several days, and the absence of a guaranteed precision outcome. Therefore, the second and recommended option is to utilize the existing models encapsulated within the Jetson Software (SDK). These models have established their reliability through their high accuracy levels, while offering a seamless integration between the Nvidia Jetson platform and the ZED camera. This approach ensures not only a pragmatic and efficient solution but also a robust one that aligns with the initial objective of providing valuable assistance to individuals with visual impairments. An image providing insight on how the model could work is shown below in Figure 4.4.

Figure 4.4: A visual representation of the Jetson's Object Detection.

CHAPTER 5

Conclusions

The main intent of the project was to explore the idea and development of an innovative assistive system that uses smart glasses and a new machine learning algorithm to aid visually impaired individuals in navigating their environment safely. Also, by using the technologies at hand, such as Jetson computers and Zed cameras, make a portable and efficent solution that ensures confort and security among visually impaired. Despite encountering challenges in the implementation of tensor regression networks due to their unknow development in the field, the project successfully proposed the usage of an alternative approaches using pretrained CNN models that are also realiable. This alternative solution maintains the seamless integration of hardware components while ensuring efficient object detection and timely guidance for users.

5.1 Further Work

The project presented a successful initiation into the idea of using tensor regression models into object detection architectures, this highlights the idea of a greater effort and future work to refine and optimize the use of tensor regression layers can be explored within any neural network model. This could lead to a tensor regression model specifically made for assisting devices in order to ensure a better life quality for those in need. This may involve collaborating with experts in both machine learning and computer vision to address the specific challenges associated with tensor regression integration, and fine tuning, as a tensor regression dropout layer can also be applied to this type of models to make the model resistant to noise and more robust to misclassification [27]. Also, the exploration of the idea to use more of the tools that ZED and Jetson already offer, in order to improve the system. Some examples of this integration are the use of camera's positional tracking to replace a GPS system, in order to memorize common passages and improve obstacle avoidance algorithms. And finally, a further exploration on the integration and usage of audio and video recording and analysis to not only improve the range of detection of the system, but also to determine where an obstacle may be coming from.

Bibliography

[1] WHO, "Blindness and vision impairment." https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, October 2022.

[2] W. Jeamwatthanachai, M. Wald, and G. Wills, "Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings," *British Journal of Visual Impairment*, vol. 37, no. 2, pp. 140–153, 2019.

[3] A. Bhowmick and S. M. Hazarika, "An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends - journal on multimodal user interfaces," 1 2017.

[4] M. P. de Freitas, V. A. Piai, R. H. Farias, A. M. R. Fernandes, A. G. de Moraes Rossetto, and V. R. Q. Leithardt, "Artificial intelligence of things applied to assistive technology: A systematic literature review," *Sensors (Basel)*, vol. 22, p. 8531, Nov. 2022.

[5] R. Manduchi and S. H. Kurniawan, "Mobility-related accidents experienced by people with visual impairment," *American Educational Research Journal*, vol. 4, no. 2, 2011.

[6] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, 2019.

[7] R. VelÃ¡zquez, "Wearable assistive devices for the blind," *Lecture Notes in Electrical Engineering*, vol. 75, pp. 331–349, 10 2010.

[8] W.-J. Chang, L.-B. Chen, C.-Y. Sie, and C.-H. Yang, "An artificial intelligence edge computing-based assistive system for visually impaired pedestrian safety at zebra crossings," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 1, pp. 3–11, 2021.

[9] S. Bhatlawande, M. Mahadevappa, J. Mukherjee, M. Biswas, D. Das, and S. Gupta, "Design, development, and clinical evaluation of the electronic mobility cane for vision rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 6, pp. 1148–1159, 2014.

[10] L. Ran, S. Helal, and S. Moore, "Drishti: an integrated indoor/outdoor blind navigation system and service," *Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the*, pp. 23–30, 2004.

[11] S. Khusro, B. Shah, I. Khan, and S. Rahman, "Haptic feedback to assist blind people in indoor environment using vibration patterns," *Sensors*, vol. 22, p. 361, jan 2022.

[12] D. I. Ahlmark, "Haptic navigation aids for the visually impaired." https://api.semanticscholar.org/CorpusID:69469041, 2016.

[13] P. R. Bhole, "Gps based voice navigation system for blind people," *International Journal for Research in Applied Science and Engineering Technology*, vol. 7, pp. 3587–3590, apr 2019.

[14] K. Singh, M. Vashisht, Jyoti, I. N. Saxena, H. Tyagi, and D. Saxena, "Navigation system for blind people using gps & gsm techniques." https://api.semanticscholar.org/CorpusID:53391410, 2017.

[15] R. Marukatat, P. Manaspaibool, B. Khaiprapay, and P. Plienjai, "Gps navigator for blind walking in a campus," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 4, pp. 1481–1484, 2010.

[16] K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 2, pp. 172–182, 2018.

[17] J.-H. Kim, S.-K. Kim, T.-M. Lee, Y.-J. Lim, and J. Lim, "Smart glasses using deep learning and stereo camera," in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pp. 294–295, 2019.

[18] P.-J. Duh, Y.-C. Sung, L.-Y. F. Chiang, Y.-J. Chang, and K.-W. Chen, "V-eye: A vision-based navigation system for the visually impaired," *IEEE Transactions on Multimedia*, vol. 23, pp. 1567–1580, 2021.

[19] M. M. Islam, M. S. Sadi, and T. BrÃ¤unl, "Automated walking guide to enhance the mobility of visually impaired people," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 485–496, 2020.

[20] T. Shah and S. Parshionikar, "Efficient portable camera based text to speech converter for blind person," *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 353–358, 2019.

[21] M. A. Khan, P. Paul, M. Rashid, M. Hossain, and M. A. R. Ahad, "An ai-based visual aid with integrated reading assistant for the completely blind," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 6, pp. 507–517, 2020.

[22] H. Murad, N. I. Tripto, and M. E. Ali, "Developing a bangla currency recognizer for visually impaired people," *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, 2019.

[23] R. Parlouar, F. Dramas, M. J.-M. Mace, and C. Jouffrais, "Assistive device for the blind based on object recognition: an application to identify currency bills." https://api.semanticscholar.org/CorpusID:8448712, 2009.

[24] B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," *IEEE Internet of Things Journal*, vol. 6, pp. 1375–1383, 2019.

[25] F. Ashiq, M. Asif, M. Ahmad, S. Zafar, K. Massod, T. Mahmood, M. Mahmood, and I. Lee, "Cnn-based object recognition and tracking system to assist visually impaired people," *IEEE Access*, vol. 10, pp. 1–1, 01 2022.

[26] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning," 06 2016.

[27] J. Kossaifi, Z. C. Lipton, A. Kolbeinsson, A. Khanna, T. Furlanello, and A. Anandkumar, "Tensor regression networks," 2020.

[28] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, pp. 455–500, aug 2009.

[29] A. Williams, "Notes on tensor decompositions." https://alexhwilliams.info/pdf/cpd_notes_janelia_2016.pdf, 2016.

[30] Y. Si, Y. Zhang, and G. Li, "An efficient tensor regression for high-dimensional data," 2022.

[31] S. Labs, "Zed mini camera and sdk overview." https://cdn2.stereolabs.com/assets/datasheets/zed-mini-camera-datasheet.pdf, 2018.

[32] Nvidia, "Jetson orin nano developer kit carrier board." https://bit.ly/3DWidrQ, 2023.

[33] J. Kossaifi, Z. C. Lipton, A. Kolbeinsson, and A. Khanna, "Tensor regression networks," *Journal of Machine Learning Research*, 2020.