Real-Time Assistive Solution for Accessible Navigation: A User-Centered Approach

Team AIbRT

Meron Abera, Dania Akil, Raghav Chakravarthy, Jaime Coreas-Parada, Brandon Fox, Kaz Henderson, Prabhat Jain, Alan Jiang, Juno Song

Mentor: Dr. Mohammad Nayeem Teli

Librarian: Zaida Diaz

Honor Pledge:

We pledge on our honor that we have not given or received any unauthorized assistance on this assignment.

Abstract

Team AlbRT seeks to help people with visual impairments navigate the world in a cost-effective and user-friendly way. To do this, we will first interview people with visual impairments to gain insight into their specific needs and challenges. With this feedback, we will develop a pair of glasses equipped with an RGB-D camera and LiDAR sensor that will help them "see". These will compute the most dangerous objects to the user, giving real-time audio feedback which the user can then react to. This research aims to create an accessible and cost-effective device to address the gap in navigation support for individuals with visual impairments.

Introduction

According to the World Health Organization (WHO), approximately 2.2 billion people are impacted by some form of visual impairment, with 39 million of those being blind (World Health Organization, 2019). In a world where sight is often taken for granted, visual impairments such as blindness affect how individuals navigate their surroundings. Daily life for people whose eyesight is very limited, or who lack any sight at all, presents many challenges. Two examples of this are traversing unfamiliar environments and simply detecting what objects surround them (Kells, 2001). Despite advancements in assistive technologies, people with visual impairments still face the barrier to full independence in our predominantly sight-driven world.

Canes and guide dogs are already popular aids being utilized by people with visual impairments. However, they fail to provide detailed information in more dynamic scenarios such as on busy roads and staircases (Hersh, 2022). Therefore, providing innovative real-time solutions that empower people with visual impairments remains critical.

Research Problem

Research has shown that many visual aids on the market are either incredibly expensive or highly inaccurate, not being able to adapt to more complex environments (Nguyen et al., 2021). In a survey of 30 participants with significant visual impairments, 21 of them reported avoiding places such as airports, museums, and universities due to the physical challenges associated with visual impairment (Jeamwatthanachai et al., 2019). This highlights a broader issue: people with visual impairments frequently face difficulties and feel less confident

navigating through daily life, especially in urban settings with numerous pedestrians, vehicles, and varied terrains.

Despite advancements in sensor-based technologies, many solutions fall short in providing the comprehensive environmental information necessary for a fully-integrated navigational aid. The advent of artificial intelligence has rapidly accelerated the development of visual impairment technologies. A variety of machine learning solutions that detect and classify objects have been developed over the last few years (Jiaji Wang et al. 2023). However, there are gaps in current hardware with AI integration. Shortcomings like low battery life and high costs (Cohen, 2023) highlight the need for a further refined device that ideally combines real-time object detection with a more user-friendly and accessible interface.

Objective of paper

Now that we've established what kind of problems people with significant visual impairments face and the shortcomings of current assistive technologies, the aim of this paper is to build a device that complements the other senses of a person with visual impairments, helping them navigate the world through real-time auditory feedback. The goal is to turn visual information into auditory feedback by leveraging embedded-systems, leaving our users with real-time descriptions of their environments. We will do this by employing panoptic segmentation to provide detailed, contextual information about objects and spaces, helping bridge the gap between static methods of current technologies and the dynamic needs of many people with visual impairments. We propose an AI-based object detection system that utilizes panoptic segmentation to analyze real-time visual data, converting it into audio cues for the user.

Altogether, the system will consist of a pair of camera-mounted glasses connected to a pair of bone conduction speakers and a Raspberry Pi, providing our users with a mostly non-intrusive solution to navigating their surroundings. The Raspberry Pi will perform all the operations needed for the AI-powered obstacle/object detection and will be located on the user's body, preferably in their pocket at the time of use. By combining modern computer vision techniques and intuitive auditory feedback, this device aims to surpass existing technologies in terms of performance, pricing, and ease of use. This solution will give users clear, detailed information about their environments, offering a more comprehensive navigation experience. A few questions we hope to answer with our device include: What design features are most valued by people with a significant lack of eyesight in assistive technologies? What percent of users can navigate a set of challenges with our assistive aid, that couldn't without it? And lastly, what kind of challenges will our real-time auditory system face, and how could we minimize those challenges?

Broader Impact

The development of this AI-based assistive technology has the potential to give those with visual impairments greater independence in everyday life by giving them an alternative aid that would allow them to reduce reliance on traditional aids and human assistance. Even beyond individual empowerment, this research aims to contribute to the broader field of assistive technology. By showcasing the feasibility of advanced AI techniques in cost effective hardware, this project aims to pave the way for further developments in assistive devices, while also

promoting inclusivity in technology design. In short, this research focuses on improving accessibility in public spaces, for individuals with visual impairments.

Literature Review

To develop an assistive device for those with visual impairments, it is crucial we understand how different levels of visual impairments differ in their assistive technologies, how they perceive the world, as well as the current state of both hardware and software technologies that could be used to improve their navigation experience. This literature review aims to explore existing research relative to these areas, laying the foundation for our solution.

This review is divided into three sections, the first of which focuses on how those with visual impairments interact with and perceive their environment. Understanding their perception comes before anything else, as it allows us to create intuitive technologies that align with their natural methods of interaction. The second section focuses on hardware solutions existing in other assistive devices. We will assess various sensors and embedded systems aimed at aiding those with visual impairments, discussing their strengths and weaknesses in real-world applications. The third section covers advancements in software, specifically computer vision algorithms, AI, and some other machine learning techniques. This section highlights the importance of panoptic segmentation, a way of detecting and classifying distinct objects, which will enhance spatial awareness for those with visual impairments.

By conducting this literature review, our goal is to identify key insights from previous studies that could help guide our design and implementation of an AI-powered device, ensuring it meets the expectations of our users.

How people with visual impairments perceive the world

In order to understand how to create an effective visual aid, it is crucial to understand how those with visual impairment perceive the world. There are different levels of visual impairments which would require a visual aid device. Furthermore, there are different stages of life in which the visual impairments may have been developed. To create a device that would be effective for each individual of the visually impaired community, no matter the level of the visual impairment or period of development, understanding and applying the collective needs of individuals who are visually impaired is a priority. As a result of the absence or impairment of one sense, the other senses of individuals who are visually impaired tend to be heightened, with hearing being particularly significant. In fact, upon interviewing a group of participants between the ages of 29 and 53 who had different levels of blindness, Kells (2001) found that most participants referenced their hearing as a main sense. When asked if they could typically notice whether or not another person or obstacle was in their way, most participants shared that they found it difficult to explain exactly how they could sense this presence, indicating that the ways in which they perceive the environment and world around them is just as natural as the ability to physically see. Most participants did point to the hearing sense as a major means of perception. However, many participants noted that in order to navigate areas and get a feel for their surroundings and environment, it often has to be done at a slower pace. Participants added that there are many environmental factors that make perception and sensing the surrounding environment more difficult, including noise, wind, and snow (Kells, 2001). These are all important factors to consider when creating a visual aid that can be used to navigate day to day situations, even in the face of possible setbacks, unexpected obstacles, or environmental changes.

Humans use a process called the Mirror System in order to learn and develop behaviors by observing others, mostly through vision or what humans physically observe occurring in front of them. According to Dr. Rajmohan and Dr. Mohandas, the Mirror System is a group of specialized neurons in the brain that are activated upon watching another individual perform a behavior. The role of these neurons is to then turn these observations into information and knowledge inside the brain, leading to the mirror of these behaviors (Rajmohan & Mohandas, 2007). The Mirror System is very closely linked to social cognition, specifically emotion processing, face recognition, and memory. The activation of neurons in the mirror system is essential to humans being able to recognize and even anticipate the actions of others, making the understanding of how this system works extremely important (Rajmohan & Mohandas, 2007). In a paper investigating whether or not this Mirror System functions similarly for those who are blind, the findings were that individuals who were blind performed essentially the same as individuals without any visual impairments on a brain scan to see where different areas in the brain light up to indicate the Mirror System and its neurons at work, but this was mostly when listening to familiar action sounds. The findings of this paper suggest that the brain's Mirror System can adapt to various sensory inputs, highlighting the potential for developing a visual aid that is auditory based (Ricciardi et al., 2009). Understanding that auditory inputs can activate the mirror neurons in the brain provides a helpful insight into how to design an aid that will allow the brain to process sounds in order to understand the next action.

Navigation plays an important role in day to day life and is a major reason why an effective visual aid is necessary for individuals who have visual impairments. Individuals with sight will often look at a map before going to a new place in order to get a mental representation of their intended route and plan a sequence of events of the necessary directions to take.

Guerreiro et al. (2017) aimed to investigate if that sequential representation was not only possible, but useful for blind individuals as well. In order to do so they created an app that made use of various audio effects to simulate step-by-step walking scenarios of a route. The findings indicated that participants were able to maintain a clear mental representation of the route through the audio simulation, which reinforced the hearing sense as a means of having strong spatial awareness and navigation abilities for individuals with blindness or visual impairments (Guerreiro et al., 2017).

Blindness can occur at various levels and at different stages of life. It can be developed later in life or can be congenital. When visual impairments and blindness are developed early on in life, the "systematic large-scale reorganization of whole brain cortical-thickness networks" is triggered and can differ within various regions of the brain (Hasson et al., 2016). When blindness or visual impairments are developed earlier in life, the brain can rewire itself to adapt through the use of other senses, often heightening the senses (Hasson et al., 2016). Both congenitally blind individuals and individuals with acquired blindness were found to be heavily influenced by tactile sensory channels of the brain, which suggests that "there is a perceptual symbol representation in the conceptual representation of the blind population" (Shen et al., 2022). Understanding these commonalities of the conceptualization and perception of the world between individuals with congenital blindness and acquired blindness is important in understanding the needs that the device will have to address in order to be as effective as possible.

Hardware solutions

Hardware solutions for assisting visually impaired individuals are built from a range of components including sensors, cameras, and processing systems. In order to create viable navigational technology for people with visual impairments, it is crucial to understand current hardware products in the market and the main components used to create such products, examining their performance and limitations.

Recent advancements in assistive technologies have led to the development of innovative devices, including smart canes and smart glasses, which significantly enhance the quality of life for individuals with visual impairments. Among the prominent offerings in this field are Envision AI and OrCam, which provide alternative solutions beyond traditional smart canes (Envision, 2020).

Envision AI is at the forefront of smart eyewear technology, creating products that resemble popular frames such as Ray-Ban Meta and Amazon Echo. These smart glasses incorporate assistive features, including voice commands and touch panels, enabling users to take photographs, inquire about captured images, interact with AI models like ChatGPT, and share their surroundings via video calls (Envision, 2020). Such functionalities not only cater to individuals with visual impairments but also appeal to a broader audience, enhancing the utility and desirability of the product.

In contrast, OrCam offers an innovative solution for users who prefer to retain their existing eyewear. Rather than requiring a complete replacement, OrCam's attachment seamlessly integrates with conventional glasses, providing similar capabilities to dedicated smart glasses (Cohen, 2023). However, despite the promising features of both Envision AI and OrCam, current iterations of these devices are hindered by several design limitations, including weight, cost, as

well as a lack of essential features such as customizability, obstacle detection, and real-time processing. These shortcomings contribute significantly to the low adoption rates among individuals with visual impairments.

Chien and Snyder (1975) conducted research on the hardware requirements essential for effective visual image processing. This paper highlighted the importance of balancing speed, resolution, and dynamic range in the selection of imaging devices. In contexts involving obstacle detection, they noted that speed is very important and dynamic range can be diminished in favor of resolution. The paper also emphasizes the significance of minimizing noise within the imaging system. Many existing devices, particularly those utilizing outdated imaging hardware, fail to adequately address issues such as shot noise, thermal noise, and amplifier noise, resulting in suboptimal signal-to-noise ratios. As noted by Chien and Snyder (1975), these devices are unable to detect the shot noise of imaging devices, thermal noise in the load resistor, as well as the amplifier noise, while also lacking capabilities to reduce noise for a manageable signal-to-noise ratio.

To enhance image quality and processing capabilities, gamma correction is one potential solution for mitigating image processing challenges in these devices. Although linearity is generally not a major concern, applying a logarithmic grayscale transformation can improve the brightness and clarity of images, thereby facilitating a more accurate representation of real-world environments (Chien and Snyder, 1975). This approach aligns closely with the visual processing capabilities of the human eye, making it an attractive option for integration into computer image processing systems, particularly for applications aimed at assisting individuals with visual impairments.

Cameras and Sensors

In examining the current landscape of assistive technologies, it is crucial to recognize that their effectiveness is heavily reliant on the quality and integration of underlying hardware components. Sensors and cameras are the fundamental parts of most navigational aids for visually impaired people, serving as the primary means for gathering environmental data. The choice of these components greatly influences system performance, with the various options offering their own advantages and limitations when it comes to accuracy, environmental understanding, and cost-effectiveness. Balancing these factors is critical in creating solutions that are both practical and accessible. Therefore, a thorough evaluation of these hardware elements is essential for the development of assistive devices.

RGB-D (Red, Green, Blue, Depth) cameras, which are among the more popular technologies used for navigation, provide data on both color and depth in real-time. This dual functionality allows three-dimensional representations to be created, with depth data enabling for identifying the proximity of objects. Cameras of this type, such as Microsoft Kinect, have attracted attention for its ability in obstacle detection, working well in indoor and outdoor settings (Tychola et al., 2022). Despite this, without the various algorithmic programs that enhance their performance, RGB-D cameras are otherwise sensitive to texture and lighting conditions, leading to inaccuracies when it comes to detecting objects with complex textures or in varying lighting conditions. Furthermore, RGB-D cameras also bring about challenges with memory consumption and processing power, when used for detecting more complex objects (Abidi et al., 2024).

In contrast, infrared and ultrasonic sensors, although less detailed, tend to excel in environments where RGB-D cameras may fail. Rather than relying on light, infrared sensors use

heat signatures, allowing users to detect people and objects in darker lighting conditions. This ability is especially useful for nighttime navigation or dimly lit environments such as underground or poorly lit areas. Along with operability in the dark, these sensors are affordable and discreet. However they also have their own setbacks, most prominently issues with shorter range and possible interference from other infrared sources, such as sunlight or fluorescent light (Jafri et al., 2017). Ultrasonic sensors are another common component in assistive navigation devices, operating by emitting sound waves that reflect off nearby objects to provide information about certain objects. Such sensors are effective at close-proximity detection, which make them great for short-range navigation and detecting nearby obstructions. In contrast to infrared sensors and RGB-D cameras, their detection performance remains stable in various lighting conditions, including complete darkness. They are also relatively inexpensive, but tend to fall short when it comes to level of detail, only being able to detect objects within a specific range (Abidi et al., 2024).

Given the strengths and weaknesses of individual sensors and cameras, combining data from multiple sensor types can reduce the shortcomings of each individual technology. This approach of fusing sensors aligns with broader research into SLAM (simultaneous localization and mapping) technologies (Chen et al., 2022). Chen et al. (2022) offers a detailed exploration of how fusing data from multiple sensors can enhance performance. This paper has a specific focus on fusing LiDAR which excels in distance measurement and 3D spatial mapping, with Visual SLAM which make use of RGB-D cameras to provide rich texture and color information. In combining them, successful results were produced in which there was higher mapping precision and minimized errors (Chen et al., 2022). With success in incorporating multiple sensors, a more

effective and adaptive system can be created, allowing visually impaired individuals to navigate various environments more safely.

Abidi et al. (2024) has also found the benefits of integrating cameras and sensors, including enhanced perception, robustness, and safety. For example, utilizing RGB-D cameras with tactile sensors allows for combined visual, touch, and contact information, enhancing object recognition and feedback. While camera-sensor fusion has many advantages and applied benefits, Abidi et al. (2024) and Chen et al. (2022) both emphasize the drawbacks of such technology when it comes to processing. Implementing multiple sensors complicates system design as complex algorithms are necessary to process and fuse data. In addition, factors such as increased cost, increased maintenance due to system complexity, and calibration and synchronization are also challenges to be considered (Adidi et al., 2024; Chen et al., 2024).

Overall, each sensor has its own strengths and drawbacks, but the potential of sensor fusion is promising in overcoming individual limitations. By integrating data from multiple cameras and sensors, the richness of environmental information is enhanced, allowing for more comprehensive situational awareness. Research on multi-sensor fusion algorithms such as those conducted by Chen et al. (2022) and Liu et al. (2022) have contributed to this integration, enhancing accuracy and reducing complexity in navigational systems. Such research highlights the growing potential of sensor fusion to improve navigation performance across a variety of applications. However, while there are many possible sensor and camera combinations for navigation, the choice of fusion method depends on factors such as the target application, available resources, and desired balance between accuracy, efficiency, and cost. Beyond this, the successful integration of multiple sensors with multi-sensor fusion algorithms is also highly dependent on the hardware's ability to process this data well and in real time. Thus, compact,

effective processing solutions must also be explored to ensure the functionality and portability of such fusion technology.

Processing hardware

Creating a device that can efficiently process the large amount of data gathered by cameras and sensors in real time is a significant challenge. In order to create a device that can effectively be used by the visually impaired, it is important to use hardware that has proper processing power so that it is powerful enough to handle the necessary complex operations and real-time processing. At the same time, however, the processing hardware needs to be compact enough to allow for a non-invasive, portable and comfortable solution for the user. Additionally, it is essential to find hardware which minimizes the amount of energy used so that it can last longer on a single charge, allowing it to be used for longer periods. Finding an optimal hardware solution that can achieve all of these tasks while remaining at a relatively low cost is important in creating the best product possible.

There are four types of processors commonly used to run machine learning algorithms: Graphic Processing Units (GPU), Application Specific Integrated Circuits (ASIC), Field Programmable Gate Arrays (FPGA), and Microcontrollers (Diab et al., 2022). These processors were evaluated based on their performance, energy usage, physical space, price, and flexibility in usage. Out of these options, GPUs have the highest computational power and fastest calculation time. The main drawback to using a GPU for this application is the large size and high amount of energy needed, resulting in poor battery life of the device. The next option is the ASIC, which stands out for its ability to be designed to complete a certain task, leading to good performance in the task. This results in this processor being very energy and space efficient as well. However,

the development and production of ASICs can be very expensive, as well as its lack of flexibility. This is because once it is produced, it can only be used for what it was designed for (Diab et al., 2022). This means any change to the product would not be as effective for the ASIC as it would for the other options. The FPGA is relatively energy and space efficient, being much more efficient in both of these areas than the GPU, but less than the ASIC. The performance of the FPGA has slower performance than the previously mentioned processors, but it comes at a reasonable price. Finally, the Microcontroller, which is second most efficient in regards to both energy and space. The main drawback of a Microcontroller is that the performance is highly dependent on the specific Microcontroller and the task that it is completing (Diab et al., 2022).

The best processor for our project will be the processor that has the needed computational power to run the machine learning algorithms and is of correct size, price, and energy efficiency. Although the GPU and ASIC are both processors which have strong advantages over the others in specific aspects, they both have certain limitations that make them less than optimal choices for our task, with the GPU being too large and energy inefficient and the ASIC being too expensive. The two alternate options, the FPGA and Microcontroller, are processors which are acceptable in each of the necessary categories. Since Microcontrollers are more energy efficient, many people consider Microcontrollers to be the most appealing option (Diab et al., 2022). A common concept that is used to connect these processors, either the Microcontroller of FPGA, is called TinyML, or Machine Learning TinyML (Han & Siebert, 2022). This concept focuses on low power consumption and low cost whilst operating with the necessary performance. Considering the hardware limitations previously mentioned, TinyML provides an opportunity to balance all the performance and efficiency needed in these edge devices. Image classification

was a use case for this concept estimated to be 17.39% of all uses for TinyML, demonstrating how this can be connected and used for our project (Han & Siebert, 2022).

As technology advances, processing units will continue to shrink in size while maintaining or increasing performance and energy/spatial efficiency. However, a large part of this efficiency depends not only upon the hardware of our product, but also on the software components incorporated into the device.

Software Solutions with Panoptic Segmentation

An important consideration when dealing with large continuous image data, is to figure out how to map important scenes within each picture. To do this, we use segmentation, which aims to break the image down into smaller sub-parts that aid in further processing/interpretation of the parts of the image. There are three types of segmentation: semantic segmentation, instance segmentation, and panoptic segmentation (Abidi et al., 2024).

Semantic segmentation identifies different objects in a scene but does not distinguish between different items of the same type. This is great for gaining a general understanding of the scene and gives an outlook into what may be going on. Convolutional Neural Networks are used to implement semantic segmentation, as these algorithms look for general similarities and trends within images through processing (Long et al. 2015).

Instance segmentation identifies each instance of any object. This is necessary as it can help provide context to an image based on the number of recurring elements. It fills the gaps of semantic segmentation as it gives an idea of the magnitude/scale of an item's presence in the scene. The nature of this algorithm helps enable special decision making based on the number of

instances of a class, which helps with scenarios that deal with real-time classification, like highway traffic (Hafiz et al., 2020).

Panoptic is a combination of the above two methods, as it gives each item type one label, but each instance is uniquely named. This is the preferred method, as it allows for a holistic view of the scene, and more importantly, an accurate representation of the surroundings. Segmentation, especially panoptic segmentation, is notoriously data heavy and computationally intensive. There has been work done to speed up these segmentation algorithms through the implementation of self-supervised learning, where clustering algorithms are used to create pseudo labels of the data (Verma, 2023). This reduces the dataset size necessary, and allows for accurate grouping of similar data points and improves efficiency in the algorithm runtime, which is vital for computationally heavy projects such as this.

In order to maximize the precision of panoptic segmentation (which is a meld between semantic and instance segmentation techniques), there are 2 separate models used to segment the image at a time. Since semantic segmentation is utilized for its ability to classify different item types, a Fully Convolutional Network is used to assess the scene pixel by pixel, and to count the amount of items with instance segmentation, a Mask R-CNN is used (Elharrouss et al., 2021). After these separate models are run, panoptic segmentation uses a heuristic function, which aims to maximize the accuracy of detection, and assign precedence based on the item being tracked. For more general parts of the image (sky, land, water, etc.), semantic takes precedence, while for objects/people, instance segmentation model takes precedence. Through continual testing on datasets, such as the COCO Panoptic Dataset, the heuristic function accuracy increases and becomes more fine tuned for use in real-time scenarios (Elharrouss et al., 2021).

Depth Perception

Another extremely important factor in identifying objects is the priority of the objects, letting our product inform the user of the highest priority or most dangerous objects first. The main factors in determining priority of objects are the type of object and the distance the object is from the camera, otherwise known as depth perception. Closer objects are a higher priority, as they have a higher chance of harming the user. We will be talking about two main ways of finding the distance of an object, which are LiDAR and single-image depth perception.

Light detection and ranging (LiDAR) sensors send laser light out and measure how long the light takes to come back to the sensor. Doing this for all objects, the sensors generate a distance map for the objects in view. Within LiDAR, there are three main subsections: 1D, which provides a view in front of the camera, similar to an ultrasonic sensor; 2D, which provides a 360-degree horizontal view; and 3D, which provides both a 360 horizontal view and a nonzero vertical view. However, the costs go up the more dimensions LiDAR scans, with 3D LiDAR sensors costing hundreds of dollars. Therefore, 2D and 3D LiDARs are likely to be unviable for our research, as we want to find a cost-effective solution. 1D LiDARs, though, are more cost-effective and would work well with our hardware solution, as we are looking for a light and small device (Patel et al., 2024). Then, researchers combine LiDAR data with other cameras and software such as Convolutional Neural Networks (CNN) to detect obstacles. Chitra et al. (2021) used a 1D LiDAR sensor with a normal camera attached to a belt, which identifies objects with a CNN and measures distance. However, while more cost-effective than 3D, it is still bulky and limited in the variety of objects it can detect. Additionally, as the LiDAR sensor is 1D, it has a very limited field of view. Overall though, various LiDAR designs can provide readings with accuracies of 93.1% of the environment up to one kilometer away, far beyond what we need.

They have an error margin within two centimeters, providing accurate measurements for people with visual impairments (Patel et al., 2024). However, costs of higher degree LiDARs prevent them from being viable and the low field of view of the 1D LiDAR presents challenges in detection.

Another option for depth perception is simply using a camera and estimating depth from the image, which is more cost-effective and less bulky than adding a LiDAR sensor on top of a camera. Aleotti et al. (2020) did this by first training various neural networks such as MonoDepth2 (Godard et al., 2019), PyDNet (Poggi et al., 2018), and FastDepth (Wofk et al., 2019) on datasets such as KITTI. They then tested these pre-trained neural networks, getting accuracies of up to 97%. Lastly, they put these pre-trained neural networks onto mobile devices and tested them in real-time, getting an FPS of 10 for MonoDepth2 and around 50 for PyDNet and FastDepth, much more than necessary for simple depth perception. However, the model's accuracy suffered when tested outdoors. They did not have enough outdoor image datasets. Additionally, they did not merge this depth map with object detection or test it with an external camera instead of a handheld device's camera (Aleotti et al., 2020). Overall though, this paper provides a great alternative to LiDAR that keeps costs lower while still functioning relatively well.

An alternative single-image depth perception option comes from Ranftl et al. (2022). They made a zero-shot design, meaning a design that can perform tasks it hasn't been explicitly trained on. In this case, they used a collection of datasets, including 3D movies they analyzed themselves, in their training. Although this larger dataset results in increased accuracy and a successful zero-shot design, which is extremely important to our use case, they did not run it in

real-time, making it unviable for our use case unless altered or made more efficient. In addition, it still has failure cases, such as mirrors and subtle background blurs (Ranftl et al., 2022).

Overall, LiDAR and single-image depth perception have strengths and weaknesses in relation to each other. Single-image depth perception, while more cost-effective and lighter, would be slower and more complicated than the more expensive LiDAR. However, both options have room for improvement in the context of helping people with visual impairments navigate the world, such as real-time use or wider ranges of view.

Classification

Object classification is a crucial step in segmentation, as segmentation algorithms must classify objects before assigning a label to pixels. Clustering visual data is essential for enhancing the understanding, interpretation, and manipulation of complex visual information we will collect through. It will enable better organization and improve the analytical capabilities of our product. In terms of image segmentation, there are a few clustering algorithms that are most prominent. One such clustering algorithm is called K-means clustering, an iterative algorithm that randomly assigns k centroids, which represent the centers of clusters. Each point in the image is then assigned to the nearest cluster, and the new centroid for each cluster is computed by minimizing squared Euclidean distances within the clusters. This process is repeated until the changes in each centroid fall below a given threshold, upon which the algorithm ends. K-means clustering is useful because the algorithm is simple and can deal with large amounts of data, quickly converges to accurate clusters, and its overall simplicity (Pugazhenthi & Singhai, 2014).

A fundamental issue with this technique is the random selection of centroids at the start, which affects the final result. Poor centroid selection can lead to extremely small clusters, or in some cases “dead centers” when a cluster has no members. Thus, heuristics have been developed

for the selection of centroids. Dhanachandra et al. (2015) utilized a subtractive clustering algorithm to find an optimal data point for the initial centroid based on the density of surrounding data points, which is called its potential value. A nearby centroid will reduce the potential value of a point, and this process is continued until k centroids have been designated. Another limiting factor with K-means clustering is the number of clusters necessary to properly segment an image needs to be known in advance; calculating k in real time would prove quite difficult. Pugazhenthi & Singhai (2014) posed a solution to the selection of both centroids and the K value. They utilize an algorithm by Yao et al. (2013) involving the distribution of image gray levels to determine k, and then select the centroids based on the peaks of the derivative of the image histogram.

A clustering algorithm that is commonly used for segmentation is mean-shift analysis. This technique is useful in research such as ours because it does not require a pre-determined k value (as K-means does), nor are centroids randomly selected. Thus, there are never dead spots with mean-shift. Given a distribution of pixels in our feature space, we can determine a density function for these pixels. With mean-shift, we start with a base point, and establish a circle around this point of radius r. The base point is then shifted by a weighted mean of its distances to other points within its radius. We repeat this process, and eventually the base point will converge to a mode, which will represent a centroid of a cluster. Points that converge to the same mode belong to the same cluster. The selection of the initial radius is extremely important. Gandhi et al. (2014) used both a modified Canny edge detection algorithm and circle Hough Transform to do so, concluding that Canny edge detection yielded better results.

According to Souza et al. (2016), minimum-cut clustering is another prominent algorithm used in segmentation. In this method, each pixel is a node, and there are edges

between it and adjacent pixels. These pixels have a weight based on their similarity in the feature space. Based on these weights, a cut is made across edges that splits the nodes into disjoint subsets, with the "cost of cut" being the sum of the weights of the edges that are cut. The association of each subgraph is the sum of all of its edges. Minimum-cut seeks to minimize the ratio between the cost of cut with the association in order to compute the optimized cuts. Souza et al. (2016) combined min-cut with anisotropic filtering methods to better segment the image, utilizing anisotropic diffusion to eliminate noise towards the interior of the image while preventing excessive information loss toward the boundaries. In order to compute the min-cut, the Ford-Fulkerson algorithm was utilized. An issue with the min-cut algorithm is that its time complexity is NP-complete, and thus requires a lot of computing power, even with low values such as k = 3. Kobori & Maruyama (2012) accounted for the heavy computing cost of the min-cut algorithm with an FPGA, and used the push-relabel method to achieve better results on the FPGA.

Conclusion

To conclude, developing assistive technologies for visually impaired individuals involves a multi-faceted approach that incorporates an understanding of how these individuals perceive the world, along with a strong grasp of hardware and software solutions. Research indicates that blind and visually impaired people rely on heightened hearing to navigate the world around them, suggesting aids that utilize auditory inputs can act as effective substitutes for visual information. To ensure these technologies meet their needs and preferences, ongoing engagement with visually impaired individuals through surveys and user-testing is crucial. Shifting focus to

the technological aspects, the various available hardware and software solutions must also be considered. The hardware solutions explored, such as RGB-D cameras, infrared, and ultrasonic sensors, each have their own individual strengths and weaknesses, but integrating them and fusing data has shown promise in mitigating these limitations. To account for the increased system complexity that will result from integration, processing units should also be considered. Microprocessors in particular have emerged as promising candidates for energy-efficient hardware processing. Finally, depth perception methods, including LiDAR and machine learning have shown to be crucial for enhanced object detection, despite challenges like cost, real-world accuracy, and bulkiness. Ultimately, these individual aspects including panoptic segmentation for comprehensive scene understanding, TinyML for efficient processing, and sensor integration for enhanced environmental detection, pave the way for creating more accessible and effective navigational solutions, despite their individual limitations.

Methodology

Our research is guided by key research questions that are focused on creating a more accessible and affordable assistive navigational device. We seek to understand what challenges that our target audience currently faces with current navigational solutions and explore how those current solutions can be made more accessible to the general public, while maintaining performance and prioritizing user-friendliness. To support these primary objectives, we investigate the role of different camera and sensor combinations in environmental mapping, the potential of AI to enhance obstacle detection, and the most effective way to communicate

navigational cues to the user. These important guiding questions work as a foundation to shape the direction and scope of our research, as well as the research design of our methodology.

The focus of our research contribution is to develop a new assistive device that makes navigation easier for visually impaired users. Our approach combines advanced yet cost-effective hardware and software to create a solution that is smaller, more affordable, and more user-friendly than many current options. We aim to achieve the same high-level accuracy found in more complex and expensive devices while simplifying the design to minimize bulk and reduce production costs. By developing this device, we hope to offer a practical and accessible navigational aid, addressing an important gap in the assistive technology market.

Design and Novelty

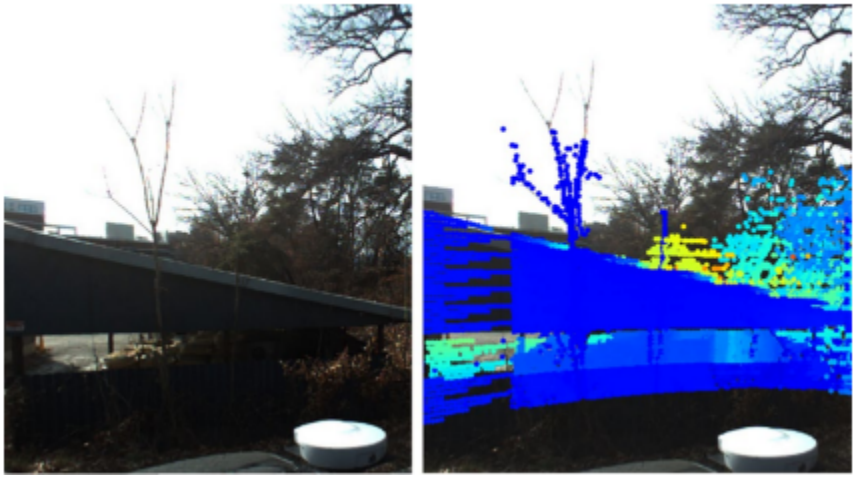
We plan to create a device that has core hardware that consists of an RGB-D camera and a LiDAR sensor, chosen to complement each other in environmental mapping and obstacle detection. The RGB-D camera captures color and depth, while the LiDAR technology provides reliable spatial mapping and depth perception in differing lighting conditions where RGB-D cameras may not perform.

Depth perception is the ability to see objects in all three dimensions, letting people perceive how far away objects are. For our device, we need depth perception to understand what priority to give objects. Objects that are closer have a higher priority, as they are more likely to interact with the user. Therefore, we need to know how close objects are so we can let the user know what is the most dangerous or highest priority. However, just a simple camera has a hard time finding the distance of objects, as the images are two dimensional. It is possible to estimate

distances, but it is less accurate. Therefore, we will use LiDAR. LiDAR, as mentioned earlier, is a type of sensor that shoots laser light out and then measures the time it takes for the laser to bounce back. The time it takes for the light to return gives the distance of the object from the sensor. This is then done for all objects in the area to get an accurate point cloud, as shown in figure 1, for all the objects in the scene. Lastly, we will combine this point cloud with our camera feed to map objects to distances, giving us the information necessary to provide audio output to the user.

Figure 1

Normal Image with Point Cloud



Note. The first image shows a scene and the second image shows the same scene with the point cloud projected on top (Javed & Kim, 2022).

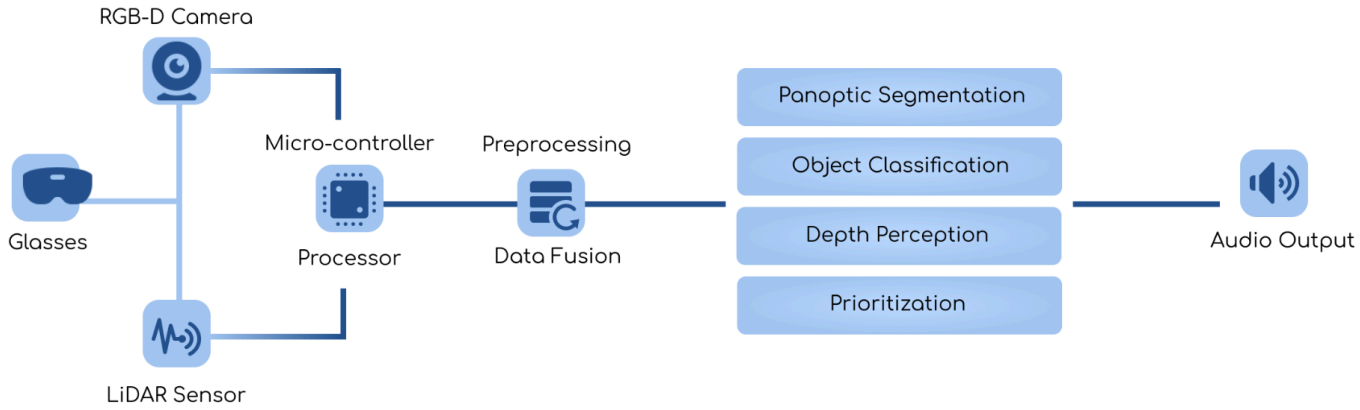
The camera-sensor fusion offers accurate environmental understanding, crucial with bringing the device on par with existing high-end navigational aids. In addition, since this system only relies on one camera and one sensor, the device can be kept lightweight, avoiding unnecessary bulk and reducing high costs often associated with multi-sensor navigation aids. By

lowering production costs and making the system more lightweight, the approach supports the development of an assistive device that offers high-functionality in a more practical and accessible form than many existing solutions.

To manage data processing demands effectively, the device will also incorporate a lightweight yet powerful processor capable of handling the data from both the camera and sensor. The processor must have the proper balance of computational power, cost, energy efficiency, spatial efficiency, and ease of use. After evaluating many processors, we ruled out Graphic Processing Units and Application-Specific Integrated Circuits: GPUs are too large and energy inefficient to be viable, while the development costs of ASICs are too extreme. This led us to considering Microcontrollers and Field Programmable Gate Arrays, which both provide the necessary computational power (Diab et al., 2022). The processing power of this hardware will allow the device to retain the accuracy, while reducing the cost of production. From here, we leaned towards Microcontrollers due to the higher energy and spatial efficiency seen in them compared to FPGAs. This energy and spatial efficiency are important, as it creates a more comfortable and viable product. Additionally, FPGAs are generally more difficult to program, which could lead to slower development times and less efficient programming. For all of these reasons, we settled upon using a Microcontroller to ensure the most viable and user-friendly product. The implementation of our device's hardware and software components is illustrated in Figure 2 below.

Figure 2

Assistive Device Full Hardware and Software Integration Flow Chart



Note. The above chart models the integration of the hardware and software components of our device, which results in an audio output to the user.

The proposed device consists of lightweight, ergonomically designed glasses with a LiDAR sensor and RGB-D camera mounted on the bridge and upper rim of the frame for optimal field-of-view. The battery will be housed in one of the arms ensuring compactness and balance, while minimizing bulk. The glasses will wirelessly connect to a separate handheld module that houses the microcontroller processor, ensuring the glasses remain lightweight and comfortable. This handheld unit will process environmental data in real-time and transmit audio feedback to the user via bone-conduction speakers embedded in the glasses arms. The setup allows the glasses to stay sleek and non-intrusive, while leveraging the processing power of the handheld unit for efficient obstacle detection and navigation assistance.

To establish the novelty and validity of our approach, we will evaluate our device's performance against key metrics including accuracy, speed, use in real-world situations, and user accessibility. To test these metrics, we will do a combination of field research and case studies. The field research will test accuracy, speed, and use in real-world situations through testing of

our device in various settings. Then, the case studies will test user accessibility and actual use by people with visual impairments, thus showing us how useful the device will actually be to them.

The accuracy of our device will be assessed by its ability to correctly detect, classify, and measure the proximity of obstacles in a variety of real-world settings. For the RGB-D Camera and LiDAR sensors, this means testing their precision and reliability in object detection and proximity when implemented with panoptic segmentation. The RGB-D camera, which combines both color and depth information, will be assessed for its ability to maintain accuracy in dynamic environments, such as environments with varying lighting conditions or low texture where its accuracy typically diminishes (Abidi et al., 2024). The LiDAR sensor should be able to detect depth in real-time with high accuracy, then prioritize what objects are most dangerous for the user based on distance and the type of object.

Once the cameras are properly calibrated, one must make sure that the software is as accurate as possible in designating danger so that it gives the best feedback to the user. This is partly done by means of panoptic segmentation, which will detail the important objects in our scene. Our testing will initially focus on verifying the accuracy of detected objects in a static picture then slowly move to dynamic scenes (videos).

The speed of our device is essential, as it is important that the user is able to get auditory feedback from the device in real time. Any delays in the transfer of information from device to user can result in a poor experience and may even endanger the user. The time which the device takes to give feedback to the user is the sum of many tasks; first, the cameras and sensors must observe the surroundings and relay this information to the processor; next, the processor must run the Machine Learning algorithm on this information; finally, the processor must output the auditory information to the user. Due to these reasons, the latency is heavily dependent upon the

computational power of the processor. Our proposed processor, the Raspberry Pi 5, is a powerful microcontroller, with 8GB of RAM. Old Raspberry Pi models with only 1GB of RAM have been able to run vision-based classification algorithms in 0.6 seconds (Htet et al., 2020). We also plan to implement a Raspberry Pi AI Hat, an accessory which can drastically improve Machine Learning performance by adding a built-in neural network accelerator. With this hardware, we believe that our algorithms will be able to run with low latency, in order to allow a more effective device.

We will also measure our contribution through rigorous user trials that assess the effectiveness and practicality of our navigation system. Testing will take the form of surveys, interviews, and field testing, where a sample of participants will use our product and provide feedback. In gathering this vital data from visually impaired users, we can evaluate how reliable our device is in real-world scenarios. Furthermore, aspects such as ease-of-use and effectiveness of feedback mechanisms can also be evaluated from this testing. This approach ensures that the device not only meets technical requirements, but also aligns with user needs, providing more qualitative insight into the validity of our contribution.

Research Design

Our research design incorporates a combination of lab research, field research, and case studies involving visually impaired users, aiming to achieve a well-rounded assessment of our devices effectiveness.

The initial phase will focus on lab research, where the RGB-D and LiDAR sensor, along with the device's processor and software components, will be tested in a regulated setting. This

allows us to systematically test the devices accuracy and reliability in a controlled environment where we can adjust specific variables and conditions. Such variables include, but are not limited to, lighting, object type, and distance, and will be modified to gain better insight into performance over varying conditions. To do this, we will first test the software and hardware separately, making sure each individual component works. Then, we will merge the parts together and test how well it does as a group. After many iterations of this process, we will have a product that works in the lab.

Following this, field research will be conducted to address the device's real-world performance in diverse environments. This is done in settings like busy streets, cluttered indoor spaces, and night time or dimly lit areas. In testing within complex and variable environments with fluctuating light, obstacle, and noise conditions, everyday challenges can be closely simulated. This field research will also rely on the surveys and interviews of our audience, to determine what specific environments are most difficult to navigate or are most commonly seen in day to day life. This aspect of research will aid in adjusting the system for adaptability, with an aim to validate that the combined RGB-D and LiDAR approach can function reliably in dynamic settings, ensuring users can navigate safely and with confidence. If needed, we will go back to lab research with our interview data and field testing, further iterating on our device to improve its performance, ease of use, and capabilities.

Finally, case studies involving visually impaired users will be conducted to gain more insight into the practicality and user-friendliness of our device. The trials will consist of specific routes that participants will be navigating, with varied conditions (e.g. indoors, narrow paths, uneven surfaces, etc.). During these case studies, users will provide feedback on various aspects of the device such as ease-of-use, comfort, and guidance ability. To ensure consistency,

guidelines will be established for each trial, including detailed instructions on how to use the device along with an initial period where users can familiarize themselves with the device and its controls. In order to obtain meaningful results, we will use a combination of observational data, surveys, and interviews. Observational data will be gathered through the trials, focusing on specific measurements such as time taken to complete routes, number of obstacles detected, and hesitation or struggle with navigation. Surveys will be given directly after each trial to capture the participants initial impression, and once all trials are complete, participants will engage in interviews where they can discuss their experiences in depth and provide specific feedback. Questions in these interviews will cover aspects such as device feedback quality, comfort, and how the device compares to other navigational aids they may have used. This combination of data and feedback will provide an in-depth understanding of the device's usability, allowing us to refine necessary aspects to best meet the needs of our users.

Given that our goal is to create a visual aid that is as effective, operative, and as helpful as possible, gaining feedback directly from individuals who are blind and who are likely to use the device is essential. We will begin the process by choosing a sample of participants who are blind and visually impaired. In order for the sample to be representative of the population, we will include participants who developed blindness or visual impairment at different stages in their lives as this may indicate different levels and areas of need. The sample will also include a wide and diverse age range of individuals ages 18 and older. Participants will be initially given a survey asking about any current device experience, main navigation needs, preferred features, and comfort and design preferences. The team will take this feedback into careful consideration throughout the design and development process. Once a prototype has been created, we will reach out to the participants and ask them to test the prototype for usability and effectiveness.

Participants will be asked for feedback on the prototype and how likely they would be to use it in their day to day lives. We will then use this feedback to continue the research and development process.

Data will also be extremely vital for training and testing our models required to make the product work. In depth processing, we will need LiDAR point cloud data to test our depth perception, as we want to make sure it works as well as possible. One such source will be the KITTI dataset, used by previous researchers such as Aleotti et al. (2020). This dataset has thousands of training point cloud data points to use, letting us test our software on a variety of sources.

Another way we will get data is through our own testing. In a controlled lab environment, we will set up various obstacles at differing distances and diverse lightning conditions to observe how accurately the device maps its surroundings. This will allow us to collect important data on depth and distance values of the sensors, which will be documented as the baseline performance metrics. In addition to collecting data on our camera and sensor, processor performance will also need to be assessed. Information on processor efficiency, power consumption, and latency are key aspects to be observed in real-time. Such data will provide insight on the processor's capacity to handle multiple data inputs, crucial in determining if it's suitable for use in our device.

After the in lab trials are completed, we will begin data collection for algorithm outputs, which will include object classification labels, depth maps, and segmentation data. This data will be collected in outdoor and non-controlled environments in order to assess the real-world application and effectiveness of the developed device.

Teams

The team will be broken up into three sub-teams that will focus on specific areas of the research and development process. There will be a team focused on software development which will develop and refine the panoptic segmentation algorithm, collaborating with the hardware team to incorporate it into the physical device. Our initial goal will be to create a very simple pipeline from our software to our hardware, not focusing on the efficiency or accuracy of our model. All we want is a system that can utilize panoptic segmentation in conjunction with the smart glasses to identify objects. The more complicated our system is initially, the harder it will be to root out issues in our pipeline when they inevitably arise. There will also be a team focused on hardware development which will test various sensors and cameras, then integrate and calibrate the hardware into a pair of glasses for testing. There will also be a team focused on the perception of individuals who are blind and visually impaired, which will work closely with the interview participants in order to ensure that feedback and insights are being clearly communicated with the rest of the team and taken into careful and thoughtful consideration. Furthermore, this team will also focus on obtaining IRB approval in order to reach out to and work with the participants of the sample. Data collection and organization will be a major priority of this sub-team.

Anticipated Results

Our anticipated results are centered around successfully developing a cost-effective assistive navigation device. In terms of functionality, we expect to achieve a high level of accuracy in obstacle detection and depth perception using RGB-D and LiDAR sensors integrated

with panoptic segmentation. We hope that we can still find a way to achieve high accuracy with less expensive and more lightweight hardware components. This also means finding a processor that is capable of handling the input and modifying software to provide a balanced solution. Currently, many solutions are either too expensive or limited, and thus are not feasible for a general audience (Patel et al., 2024). We anticipate that the compact design and affordability can make the device more accessible to a broader population, addressing the gap in available assistive technology for visually impaired individuals.

In the interview process, users are expected to provide crucial qualitative insights that quantitative data alone may not reveal. In conducting our interviews and surveys, we hope to gain a better understanding of user preferences, challenges, and expectations regarding navigational aids in real-world scenarios. Additionally, feedback from user trials can help us assess the user's perspectives on the effectiveness and clarity of the navigational cues. As each user will have differing needs, discussions around the haptic feedback mechanisms can reveal valuable suggestions for enhancing the user interface, and thus contribute to creating a usable and effective device. We hope to gain these valuable insights for optimal usability and easy integration into daily life. We aim to use this feedback and data to create a visual aid that is well tailored to the needs of individuals and that addresses areas that current devices do not.

The insights that we anticipate from the participant group would be a driving factor of a user centered approach to designing navigational aids. It is crucial that the specific needs and preferences of the individuals who are likely to use the device are prioritized and directly addressed. The team hopes to create a visual aid device that will fill the gaps in current technology by increasing the levels of usability, comfort, effective object detection, and communication.

In terms of LiDAR and depth perception, the main contribution will be maintaining high accuracy and a wide range of detectable objects with the constraints of 1D LiDAR. Most current 1D LiDAR solutions, while cheaper, are either limited in accuracy or the number of objects detectable, making them infeasible for the general audience (Patel et al., 2024). We will expand on these base solutions and thus increase the viability of our product.

Another key anticipated result is selecting a processor that maintains a sufficient balance between cost and power efficiency. As our device will be integrating data from both RGB-D and LiDAR sensors, the processor needs to efficiently handle that high-data input while maintaining speedy, real-time analysis. Current systems for navigational solutions often struggle with high processing demands, which lead to issues with latency and hinder real-time functionality, especially in portable devices (Yang et al., 2018). Processors that are capable of real-time data processing however, are often expensive or have high power consumption, making them unsuitable for cost-effective, portable solutions. Our contribution will thus be developing optimized algorithms for integrating data and panoptic segmentation software that maximize processing efficiency. In doing so we can leverage the potential of our camera and sensors to achieve accurate real-time environmental mapping without requiring an expensive, high-powered processor.

Budget

Our total expected costs amount to roughly $2,985 - 3,305. The main expenses include most of our hardware components such as $350 for 1 Ray-Ban Meta Glasses prototype, $120 for 2 Garmin LiDAR sensors, $40 for bone-conduction earbuds, $100 for 4 32-bit microcontrollers,

$10 for 1 TDC7200 sensor, $600 for 4 cameras; Software - $100 for tech stacks which we will use to build and host our application which will be used by the smart glasses to store user profile as well as any data requested by the user. Apart from the hardware costs, we are planning to spend $525 on compensations for the research and testing phases of our project which include merchandise such as Mugs and T-Shirts with the UMD and Team AlbRT logos (If permission given by UMD to include school logo), as well as a giveaway prize of $50 for survey participants. Lastly we allocated some travel and miscellaneous funds for conference, travel, or any replacements we may need, which amounted to $1500 total. For additional details, please see the table below.

The budget covers the key materials and equipment needed for prototyping, testing, and developing the initial product. It focuses on hardware components like smart glasses, sensors, and cameras as well as software/tech stack costs and travel expenses. While the full scope of the project is not detailed, this budget provides a solid foundation to get the initial development work underway.

Item	Cost/ea	Quantity	Total Cost	Purpose	Link
Care Package	$50	1	$50	Giveaway as incentive for survey	Will be comprised of multiple items yet to be chosen
UMD X AlbRT Mug	$5	20 participants	$100	For Interviewee Gifts for the Market Validation Stage	Vendor TBD

UMD X AlbRT T-Shirt	$15	25 participants	$375	For Interviewee Gifts for the Testing Stage	Cotton T-Shirt (Vendor TBD)
Ray-Ban Meta Glasses	$350	1	$350	Frame for initial prototype. Used for testing purposes and finalizing which hardware would work best.	Glasses
Garmin LiDAR Sensor	$60	2	$120		LiDAR Sensor
32-bit Microcontroller	$25	4	$100		Microcontroller
TDC7200	$10	1	$10	Could potentially be needed to measure time-intervals and convert them into digital codes.	Converter
Cameras	$70-150	4	$280-600	This cost could potentially be saved by re-using the cameras from the Meta glasses we buy. Added for better estimate	Camera

				on budget	
Bone Conduction Earbuds	$40	1	$40	Will be used for audio output in device	Earbuds
Tech Stacks	$100	1	$100	Will be used to build and run our application	
Travel	$1,000	N/A	$1,000	Includes costs for conference registration, travel, etc.	
MISC	$500	N/A	$500	Any misc costs that might arise such as hardware replacement	
TOTAL	**Low:**	**$2, 965**	**High:**	**$3, 345**	

Timeline

Our team has established a structured timeline to complete our research, prototype development, and testing within a four-year period, culminating in Spring 2027 with the presentation of our final product and research findings.

By Spring 2025, we aim to develop a fully functional software solution. This phase will involve acquiring essential hardware components, such as LiDAR sensors, and integrating the efforts of our hardware and software teams. In Fall 2025, we plan to transition to prototype development by incorporating the hardware and software systems into a cohesive product, followed by initial testing.

By Spring 2026, we intend to finalize the prototype, incorporating any necessary adjustments identified during closed testing. This phase will also include initiating market validation and engaging with potential customers for further testing and feedback. Fall 2026 will focus on the final publication of our findings and submission of our thesis.

This timeline ensures that our team remains on track to deliver a comprehensive and thoroughly tested product by Spring 2027.

Equity Impact Statement

While the purpose of our device has already been outlined - to create an affordable, compact, and user-friendly assistive navigation tool for Individuals with visual impairments- its equity impact lies in addressing accessibility barriers and promoting inclusion. Individuals with visual impairments have to navigate through complex environments every day, and while current navigation aids are helpful, they're often expensive, bulky, or lack the necessary sophistication for real-time obstacle detection (Hersh, 2022). Our solution aims to reduce these barriers by making advanced assistive technology more accessible and affordable, enabling for more independence.

First and foremost, inclusivity is a central part of our design process. We will actively seek input from individuals with visual impairments throughout the development of our device. A large portion of our design approach consists of surveys, interviews, and user trials, allowing us to involve our target audience in feedback loops. In doing so, we can effectively center user needs and take their lived experiences into account. This ensures that our device is not only functional, but equitable in addressing real-world challenges. Beyond this, we also plan on using cost-effective components that focus on affordability, ensuring that economic disparities do not prevent individuals with visual impairments from benefiting from our product. In doing so, we hope to bridge the economic barrier that often leaves individuals with visual impairments, especially those from lower incomes, without effective mobility aids.

Overall, our project not only aims to address the technical challenges faced by individuals with visual impairments, but also actively contribute to the larger goal of social equity. In incorporating an inclusive, user-based design, and reducing economic barriers, we hope to create a solution that can empower individuals with visual impairments.

References

Abidi, M. H., Siddiquee, A. N., Alkhalefah, H., & Srivastava, V. (2024). A Comprehensive Review of Navigation Systems for Visually Impaired Individuals. *Heliyon*, *10*(11), e31825. https://doi.org/10.1016/j.heliyon.2024.e31825

Aleotti, F., Zaccaroni, G., Bartolomei, L., Poggi, M., Tosi, F., & Mattoccia, S. (2020). Real-Time Single Image Depth Perception in the Wild with Handheld Devices. *Sensors*, *21*(1), 15-. https://doi.org/10.3390/s21010015

Chen, W., Zhou, C., Shang, G., Wang, X., Li, Z., Xu, C., & Hu, K. (2022). SLAM Overview: From Single Sensor to Heterogeneous Fusion. *Remote Sensing*, *14*(23), 6033. https://doi.org/10.3390/rs14236033

Chien, R., & Snyder, W. (1975). Hardware for visual image processing. *IEEE Transactions on Circuits and Systems*, *22*(6), 541–551. https://doi.org/10.1109/tcs.1975.1084080

Chitra, P., Balamurugan, V., Sumathi, M., Mathan, N., Srilatha, K., & Narmadha, R. (2021). Voice Navigation Based guiding Device for Visually Impaired People. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 911–915. https://doi.org/10.1109/ICAIS50930.2021.9395981

Cohen, R. (2023). *Experience the Power of Assistive Technology with OrCam's AI Devices*. OrCam Technologies. https://www.orcam.com/en-us/home

Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image segmentation using K -means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, *54*, 764–771. https://doi.org/10.1016/j.procs.2015.06.090

Diab, M. S., & Rodriguez-Villegas, E. (2022). *Embedded Machine Learning Using Microcontrollers in Wearable and Ambulatory Systems for Health and Care Applications: A Review*. Shibboleth authentication request. https://doi.org/10.1109/ACCESS.2022.3206782

Elharrouss, O., Al-Maadeed, S., Subramanian, N., Ottakath, N., Almaadeed, N., & Himeur, Y. (2021). Panoptic segmentation: A review. arXiv. https://doi.org/10.48550/arXiv.2111.10250

Envision - *enabling vision for visually impaired*. (2020). Letsenvision.com. https://www.letsenvision.com

Gandhi, N. J., Shah, V. J., & Kshirsagar, R. (2014). Mean shift technique for image segmentation and modified Canny edge detection algorithm for circle detection. *2014 International Conference on Communication and Signal Processing*, *15*, 246–250. https://doi.org/10.1109/iccsp.2014.6949838

Gebru, I. D., Alameda-Pineda, X., Forbes, F., & Horaud, R. (2016). EM Algorithms for Weighted-Data Clustering with Application to Audio-Visual Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(12), 2402–2415. https://doi.org/10.1109/TPAMI.2016.2522425

Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3828-3838)

Guerreiro, J., Ahmetovic, D., Kitani, K. M., & Asakawa, C. (2017). *Virtual navigation for blind people: Building Sequential Representations of the Real-World*. ACM Conferences. https://dl.acm.org/doi/abs/10.1145/3132525.3132545

Hafiz, A. M., & Bhat, G. M. (2020). A survey on instance segmentation: State of the art. International Journal of Multimedia Information Retrieval, 9(3), 171–189. https://doi.org/10.1007/s13735-020-00195-x

Han, H., & Siebert, J. (2022). TinyML: A Systematic Review and Synthesis of Existing Research. *2022 International Conference on Artificial Intelligence in Information and Communicati*on (ICAIIC), 269–274. https://doi.org/10.1109/ICAIIC54071.2022.9722636

Hasson, U., Andric, M., Atilgan, H., & Collignon, O. (2016). Congenital blindness is associated with large-scale reorganization of anatomical networks. *NeuroImage*, *128*, 362–372. https://doi.org/10.1016/j.neuroimage.2015.12.048

Hersh, M. (2022). Wearable Travel Aids for Blind and Partially Sighted People: A Review with a Focus on Design Issues. *Sensors*, *22*(14), 5454. https://doi.org/10.3390/s22145454

Htet, H. T. M., Thu, T. T., Win, A. K., & Shibata, Y. (2020). Vision-based automatic strawberry shape and size estimation and classification using Raspberry Pi. In *2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)* (pp. 360-365). IEEE

Jafri, R., Campos, R. L., Ali, S. A., & Arabnia, H. R. (2017). 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*, *6*, 443–454. https://doi.org/10.1109/access.2017.2766579

Javed, Z., & Kim, G.-W. (2022). PanoVILD: a challenging panoramic vision, inertial and LiDAR dataset for simultaneous localization and mapping. *The Journal of Supercomputing*, *78*(6), 8247–8267. doi:10.1007/s11227-021-04198-1

Jeamwatthanachai , W., Wald, M., & Wills, G. (2019). Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment*, *37*(2), 140-153. https://doi.org/10.1177/0264619619833723

Jiaji Wang, Shuihua Wang, Yudong Zhang. (2023). "Artificial intelligence for visually impaired." *Displays*, Volume 77, 2023, 102391, ISSN 0141-9382, https://doi.org/10.1016/j.displa.2023.102391

Kells, K. (2001). Ability of Blind People to Detect Obstacles in Unfamiliar Environments. *Journal of Nursing Scholarship,* 33: 153-157. https://doi.org/10.1111/j.1547-5069.2001.00153.x

Kobori, D., & Maruyama, T. (2012). An acceleration of a graph cut segmentation with FPGA. *22nd International Conference on Field Programmable Logic and Applications (FPL)*, 407–413. https://doi.org/10.1109/fpl.2012.6339137

Liu, W., Liu, Y., & Bucknall, R. (2022). Filtering based multi-sensor data fusion algorithm for a reliable unmanned surface vehicle navigation. *Journal of Marine Engineering & Technology*, *22*(2), 67–83. https://doi.org/10.1080/20464177.2022.2031558

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

Nguyen, X., Koopman, J., Genderen, M. M., Stam, H. L. M., & Boon, C. J. F. (2021). Artificial vision: the effectiveness of the OrCam in patients with advanced inherited retinal dystrophies. Acta Ophthalmologica. https://doi.org/10.1111/aos.15001

Patel, I., Kulkarni, M., & Mehendale, N. (2024). Review of sensor-driven assistive device technologies for enhancing navigation for the visually impaired. *Multimedia Tools and Applications*, *83*(17), 52171–52195. https://doi.org/10.1007/s11042-023-17552-7

Poggi, M., Aleotti, F., Tosi, F., & Mattoccia, S. (2018). Towards real-time unsupervised monocular depth estimation on CPU. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1806.11430

Pugazhenthi, A., & Singhai, J. (2014). Automatic centroids selection in K-means clustering based image segmentation. *2014 International Conference on Communication and Signal Processing*, 1279–1284. https://doi.org/10.1109/iccsp.2014.6950057

Rajmohan, V., & Mohandas, E. (2007). Mirror neuron system. *Indian Journal of Psychiatry*, *49*(1), 66–69. https://doi.org/10.4103/0019-5545.31522

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2022). Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(3), 1623–1637. https://doi.org/10.1109/TPAMI.2020.3019967

Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, J. V., Fadiga, L., & Pietrini, P. (2009). Do We Really Need Vision? How Blind People "See" the Actions of Others. Journal of Neuroscience, 29(31), 9719–9724. https://doi.org/10.1523/jneurosci.0274-09.2009

Shen, G., Wang, R., Yang, M., & Xie, J. (2022). Chinese Children with Congenital and Acquired Blindness Represent Concrete Concepts in Vertical Space through Tactile Perception. *International Journal of Environmental Research and Public Health*, *19*(17), 11055. https://doi.org/10.3390/ijerph191711055

Souza, G. B., Alves, G. M., Levada, A. L., Cruvinel, P. E., & Marana, A. N. (2016). A graph-based approach for contextual image segmentation. *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, *51*, 281–288. https://doi.org/10.1109/sibgrapi.2016.046

Tychola, K. A., Tsimperidis, I., & Papakostas, G. A. (2022). On 3D Reconstruction Using RGB-D Cameras. *Digital*, *2*(3), 401–421. https://doi.org/10.3390/digital2030022

Verma, S. (2023). Self-trained Panoptic Segmentation. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2311.10648

Wang, Y., Ahsan, U., Li, H., & Hagen, M. (2022). A comprehensive review of modern object segmentation approaches. *Foundations and Trends® in Computer Graphics and Vision*, *13*(2–3), 111–283. https://doi.org/10.1561/0600000097

Wofk, D., Ma, F., Yang, T. J., Karaman, S., & Sze, V. (2019). Fastdepth: Fast monocular depth estimation on embedded systems. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 6101-6108). IEEE. https://doi.org/10.1109/ICRA.2019.8794182

World Health Organization. (2019). *World Report on vision.* World Health Organization. https://www.who.int/publications/i/item/9789241516570

Yang, K., Wang, K., Bergasa, L., Romera, E., Hu, W., Sun, D., Sun, J., Cheng, R., Chen, T., & López, E. (2018). Unifying Terrain Awareness for the Visually Impaired through Real-Time Semantic Segmentation. Sensors, 18(5), 1506. https://doi.org/10.3390/s18051506

Yao, H., Duan, Q., Li, D., & Wang, J. (2013). An improved K-means clustering algorithm for fish image segmentation. *Mathematical and Computer Modelling*, *58*(3–4), 790–798. https://doi.org/10.1016/j.mcm.2012.12.025