ABSTRACT

Title of Thesis: A NAVIGATION AND OBJECT LOCATION DEVICE FOR THE BLIND

Team Vision: Steve Caperna, Christopher Cheng, Junghee Cho, Victoria Fan, Avishkar Luthra, Brendan O'Leary, Jansen Sheng, Andrew Sun, Lee Stearns, Roni Tessler, Paul Wong, and Jimmy Yeh.

Directed by: Professors Rama Chellappa* and Cha-Min Tang**
*Center for Automation Research
*UMD Institute for Advanced Computer Studies (UMIACS)
**Baltimore VA Medical Center
**University of Maryland School of Medicine

Team Vision's goal is to create a navigation system for the blind. To achieve this, we took a multi-pronged approach. First, through surveys, we assessed the needs of the blind community and developed a system around those needs. Then, using recent technology, we combined a global positioning system (GPS), inertial navigation unit (INU), computer vision algorithms, and audio and haptic interfaces into one system. The GPS and INU work together to provide walking directions from building to building when outdoors and the computer vision algorithms identify and locate objects such as signs and landmarks, both indoors and outdoors. The speech-based interface ties the GPS, INU, and computer vision algorithms together into an interactive audio-based navigation device. Finally, the haptic interface provides an alternative intuitive directional guidance system. The resulting system guides users to specified buildings and to important objects such as cellular telephones, wallets, or even restroom or exit signs.

A NAVIGATION AND OBJECT LOCATION DEVICE FOR THE BLIND

by

Team Vision: Steve Caperna, Christopher Cheng, Junghee Cho, Victoria Fan, Avishkar Luthra, Brendan O'Leary, Jansen Sheng, Andrew Sun, Lee Stearns, Roni Tessler, Paul Wong, and Jimmy Yeh.

Thesis submitted in partial fulfillment of the requirements of the
Gemstone Program
University of Maryland, College Park
2009

Advisory Committee:

Professors Rama Chellappa and Cha-Min Tang, Mentors

Mr. Bobby Bobo
Professor Larry Davis
Professor Ankur Srivastava
Dr. Ashok Veeraraghavan
Professor Min Wu

ACKNOWLEDGEMENTS

We would like to sincerely thank our amazing mentors, Dr. Rama Chellappa and Dr. Cha-Min Tang, for their guidance and support in this project. Without them, this project would not have been possible. We are also indebted to Aswin Sankaranarayanan, Ashok Veeraraghavan, and Pavan Turaga, who provided invaluable assistance and advice. Our thanks also go to all of the visually impaired volunteers who sat with us for interviews and system tests. Working with them was an absolute pleasure. We especially acknowledge Bobby Bobo and the Columbia Lighthouse for the Blind for all their help in coordinating volunteers, testing our system, and providing a wealth of contacts in the visually impaired community. Finally, we would like to thank our team librarians, Mr. Bob Kackley and Mr. Dan Newsome, for their help with our literature review and with locating valuable resources.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Application Programming Interface
CBIR	Content-Based Image Retrieval
GPS	Global Positioning System
INU	Inertial Navigation Unit
MIPS	Microprocessor without Interlocked Pipeline Stages
SIFT	Scale Invariant Feature Transform

Chapter 1

Introduction

Visual impairment is one of the most severe types of disabilities a person must endure and, despite numerous advancements in technology, it remains a serious problem to this day. According to the National Eye Institute, about four million Americans suffer from vision loss due to diseases of the eye such as retinitis pigmentosa, macular degeneration, and glaucoma. [1] According to a 2004 study by the National Eye Institute, the annual cost associated with vision loss in the United States is greater than $67.5 billion. One of the most frustrating aspects of visual impairment is the dependence it creates on sighted individuals for navigation and object locations. Blindness is a disability that has thus far been relatively resistant to the benefits of rehabilitation technology. An amputated limb may be replaced with a prosthetic; a paralyzed limb may regain some function with electrical stimulation; the deaf may regain some hearing through cochlear prosthesis implantation. Yet the vast majority of visually impaired individuals in this country must rely on the traditional white cane that has been used for navigation by the visually impaired community for decades. The white cane is very limited in its ability to provide navigational independence for its users. It cannot easily be used to detect obstacles above a user's waist (such as low hanging branches), nor can it detect people or objects more than a few feet away. Furthermore, the white cane cannot give specific geographical location information to its user, information that is vital for navigational independence. In addition, the visually impaired still require the assistance of sighted individuals or guide dogs to lead them to most destinations. These shortcomings drive the need for research on developing innovative navigational systems for the blind.

Many technological solutions have been proposed and implemented, but none have been widely successful in improving the mobility and lives of the visually impaired. Novel approaches that integrate guidance devices into the white cane have been proposed. Many systems (for example, [2]) exist that utilize this innovation, such as the GuideCane, a device which uses echolocation to detect objects directly in front of the cane and steer the user away from them. However, this device, and similar devices that build on the white cane, change the functionality of the cane and disrupt the personal navigation methods that the visually impaired have already developed using only the white cane. In addition, as this device can only direct users around objects on the ground, it offers very little functionality over the current white cane.

The recently developed Drishti [3] system takes the echolocation principles used for the GuideCane a step further, relying on ultrasound for indoor identification of objects at the shoulder level of the users. While the system has demonstrated success in locating objects in an indoor environment, the system is limited to only identifying where potential obstacles are and not where potential objects of interest may be. In addition, the system relies solely on GPS information for outdoor navigation, which has been shown [4] to have an error of up to 9.14 meters. This level of error is sufficient to lead any user away from a target location, such as a building, in an outdoor environment. Thus, while the Drishti system is able to warn the user of hazards and obstacles above the waistline, the limitations in other areas have prevented it from being an effective accessory for the visually impaired.

In another effort to develop navigational independence for the visually impaired, scientific advances [5] have led to the possibility of using retinal implants to restore vision. However, these methods are limited to individuals that have retained function in their retina and optic nerve, which excludes those that have lost sight due to diabetic retinopathy, retinal detachment, glaucoma, and other destructive afflictions. Further, development of such devices requires a greater understanding of how the retinal pathways represent the visual world through electrical signals. Until this knowledge is achieved, use of retinal implants remains an unlikely solution for visual impairment. Thus, technological devices remain the most promising candidates for providing the visually impaired with navigational independence.

While many previous tools have been developed to include components such as echolocation, integration with GPS, and detection of Radio Frequency Identification (RFID) tags, no system has utilized and integrated these components well enough to be accepted by the blind community as a viable navigational solution. [6] The need for a new device is apparent, as the white cane has remained the most widely used and accepted navigation tool for the visually impaired despite significant advances in science and technology.

Team Vision focuses on addressing the question: How can technology address the navigational needs of the visually impaired community on the University of Maryland, College Park campus? Surveys conducted on the visually impaired community within the College Park area have revealed a desire to retain usage of the white cane, as it has become an important aspect of navigation for the blind community. Thus, the system is designed as an accessory to be used with the white cane, granting greater navigational independence over previous systems. This is achieved through the integration of a GPS, compass, INU, real-time image analysis, and an audio-based user interface. The system is divided into two major components that focus on indoor and outdoor navigation. Like previous systems, outdoor navigation relies heavily on GPS to direct users to target destinations. However, the error associated with the use of GPS is mitigated by content-based image retrieval (CBIR) methods that use a camera feed to detect landmarks and target destinations, leading the users to them. Further, the use of cameras grant the user the ability to detect distant objects that a GPS alone cannot provide information about. Combined with the use of the white cane, independent outdoor navigation can be achieved using this system.

While GPS remains important to outdoor navigation, its limitations exclude it from use during indoor navigation. Interview subjects described a desire to be able to locate personal effects such as cell phones, wallets, and keys, which may be misplaced by the visually impaired. Accordingly, the system utilizes CBIR methods to detect objects of interest specified by the user as well as important landmarks within an indoor environment, such as exit signs, doors, and chairs. This is accomplished through the use of the scale-invariant feature transform (SIFT) algorithm that provides a robust method for matching real-time images to a database of template images. With the robustness offered by the system, users can identify their personal belongings as well as points of navigational importance, granting them independent indoor navigation.

At the heart of the system is the intuitive and user-friendly audio input-output interface that promotes interaction between the user and system. Vocal commands offer the user options to move to various destinations, locate objects, and modify the settings of the system, such as volume, rate of speech, and gender of voice. This allows users to control the system based on their own personal preferences. The interface provides directions through audio cues that center users on the path to waypoints, destinations, and objects of interest. Updating continuously, the system guides users along paths, correcting their heading until the destination is reached. When used in conjunction with the white cane, the audio interface offers feedback from the outdoor and indoor navigation components, conferring navigational independence to the user.

The selection of the College Park campus as the test bed for our system was based on practicality; it is a natural choice based on the team's intimate knowledge of the campus and the lack of navigational aids for the visually impaired community on campus. Various controlled environments around College Park provided the testing grounds for the system in real-world situations. Through our testing and experiments, the ability of the system to guide users to the entrance of a building along various paths as well as to locate and lead users to objects of interest in an indoor setting was measured for success. In each experiment, the user was able to successfully complete the trials in a timely manner. With repeated use of the systems, users were able to become more familiar with the interface and able to more quickly complete tasks.

The system is able to perform the tasks demonstrated, but is limited to the available hardware used to build the system. As advances in technology and hardware arise, the system will be able to gain further function and lead to use outside of the University of Maryland, College Park campus. The current system provides navigational aid to the visually impaired on the campus when used with the white cane. Development of the system has created a more inviting campus for the visually impaired and as innovations arise, the system can extend its use well beyond the boundaries of the campus.

Chapter 2

Literature Review

2.1 GPS

Loomis, Golledge, and Klatzky [7] investigated the use of an audio interface in a navigation system for the blind as a means of communication between the user and program. The article discusses the different approaches the authors used to incorporate the GPS system with the audio interface. There have been many advances in technology that will help improve the usefulness of GPS navigational system for the blind. These advances include improved GPS accuracy and virtual acoustic displays. The effectiveness of a GPS system was tested on blind subjects by giving them each a backpack mounted computer interfaced with a GPS receiver and a set of headphones on which were mounded a motion sensing device that follows the user's head movements. Each subject was observed and timed while going through different courses. Researchers then analyzed the effectiveness of a GPS system and acoustic virtual display, comparing the time and accuracy of a subject's performance from their course with results that would be expected from a sighted individual. The researchers set forth several possible solutions that we could implement in our final product, giving us more ideas about how to best integrate GPS and auditory user interfaces. We used their findings as a foundation of what has been done and built upon it, developing a product that is less conspicuous than a backpack computer and efficient in informing the consumer how to proceed to his/her destination.

A major concern in our research was determining whether or not GPS navigation is accurate enough to warrant its use in our system. Most commercially available GPS receivers claim an accuracy of approximately nine meters, and this type of error could easily "place a user in the center of a street instead of at the curb ramp," as stated in a study [4] conducted at Western Michigan University in 2007. This research tested the effectiveness of a GPS navigation technique called geotracking, which offers improvements to the basic BrailleNote GPS tracking software. This technique involves continuing to walk towards a destination after the system informs the user that they have arrived. The theory behind this technique is that the GPS receiver is more accurate while the person is walking than while standing still. While using the technique of geotracking, the user may have to make several "passes" to locate the waypoint, however, this method results in much greater precision in locating the destination. In their experiment, nineteen subjects were given the task of locating a 25 foot circle painted on a parking lot with and without the aid of the BrailleNote GPS system. Using the technique of geotracking, subjects were successful in locating the circle 93% of the time, as compared to 12% of the time using dead-reckoning. These results suggested that GPS navigation would indeed be helpful in our final system. We decided to use GPS technology to get our users close to their targets and computer vision for terminal navigation. The results of this study also offered us a benchmark by which to compare the effectiveness of our GPS system against a commercially available product.

Consultation with TRX, a University of Maryland based technology firm, led us to consider a technology that could improve the accuracy of our GPS unit. TRX has developed a system that incorporates inertial navigation with GPS coordinates to greatly improve the resolution of the user's position. Their system consists of an INU and a GPS receiver which send information via Bluetooth to a communication module that transmits this information to a base station. The base station is a computer where TRX software integrates gyroscope, magnetic field sensor, and accelerometer data with the GPS data. GPS has the tendency to "drift" away from the user's actual position over a period of time, as though the user was actually moving. The inertial data collected by the INU would show that this apparent movement is actually false, and it will be ignored by the software that is actively sorting through all available data. This system is designed to allow someone using the base station computer to track the locations of firefighters inside a burning building. While the concept of transmitting the data to a base station would not be ideal for our application, it would be feasible for the INU to communicate directly with the computer that our system's user will already be wearing.

2.2 Computer Vision

In 2005, Ritendra Datta, Jia Li, and James Z. Wang published a comprehensive survey [8] of the current research in CBIR, including work that had been published prior to 2000. CBIR is used to retrieve images based on a given query. The authors identify key elements to a successful CBIR system that are necessary to describe the environment using algorithmic means. The first, feature extraction, is the process of obtaining the defining features of an image and associating those features with the image. These features can be used in a subsequent process to match the image with another. Features characterized by the authors fell into three broad categories of histograms, shapes, and invariant points. Histograms of images are used to describe the differential areas of an image based on such properties as color and spatial information. Shapes are a robust and key representation of objects that are major elements in an image. Local invariants are corner points and interest points that can maintain specificity even in rotation and scaling. After extracting features from each image, the next step is to be able to retrieve that image when the same or a similar image is presented. There are many methods to approaching retrieval based on image segmentation, hierarchical grouping of features, and anchor images. However, each relies on the ability of the feature extraction method to uniquely isolate one image from the next. One method to improve the efficiency of retrieval methods discussed by the authors is the annotation of images based on their key features. This allows the algorithm to use text-based search means, which are significantly more accurate than image-based means. In addition, the authors describe the real-world requirements of an effective CBIR system that will be important to consider when designing our own implementation of CBIR methods for our final product.

David Lowe developed [9] the Scale Invariant Feature Transform (SIFT) algorithm in 1999 to generate features from an image that are invariant to translation, scaling, and rotation, and partially invariant to lighting changes. This algorithm is modeled after the responses of neurons in the inferior temporal cortex of primates that is involved in vision. Features are identified as locations in the image that are maxima or minima of a difference-of-Gaussian function. This process generates features that generally occur at points of high contrast, such as corners or changes in color. Once the features have been generated for an image, the features can then be used to represent the image and compare it with features of another image. When features from two images share the same relative spatial arrangement on the whole, it can be said that there is a match between the two images. In a preliminary experiment, a sample of 20 images had various image transformations applied to them to observe how well the features still matched. Seven different transformations were applied followed by a combination of all seven. Overall, there were good matches between images in the applied transformations. In further experiments with planar and 3D objects, the objects could be identified using the SIFT algorithm even when the objects were transformed or occluded by another object. This worked for multiple objects of interest within the same image. However, one problem noted by the author is the high dimensionality and complexity involved in this system that could generate significant run-times if a large database were used. Given the robustness of the algorithm, it is well suited for application in our proposed navigation system, though, in order to create a successful system, we would have to come up with a work-around for the large database problem.

Stefan Zickler and Manuela M. Veloso of Carnegie Mellon University designed an experiment [10] to test the efficacy of using a PCA-SIFT algorithm in combination with a clustered voting scheme for object recognition of real-time video. The researchers developed this protocol in order to further detection of objects in humanoid robot vision systems that introduce confounding variables such as perspective changes, occlusion, and motion blurring. Scale-Invariant Feature Transform is currently one of the most robust algorithms for detecting objects in still images, due to its ability to account for changes in scale, rotation, perspective, and lighting. Since the number of SIFT features can be large, it is advantageous to include Principle Component Analysis (PCA) as a preliminary step to reduce this factor. The algorithm developed by the researchers called for two main stages, the training and recognition stages. In the training stage, the PCA-SIFT algorithm is run on a training video of the object from various perspectives to develop PCA-SIFT keypoints that are to be used later in the object recognition stage. The recognition stage takes real-time video stream and applies the PCA-SIFT algorithm to find keypoints and conducts a nearest neighbor lookup to identify keypoint matches. The researchers separated the experiment into two parts, one with data obtained from a hand-held camera and the other with data obtained from a moving SONY QRIO humanoid robot. In both scenes, the algorithm identifies one or more objects from a video stream in which the objects are in different positions and orientations and sometimes partially occluded by another object in the scene. Each feature is identified in the scenes and, using a voting scheme, the center of each object of interest is localized. Only those objects that presented a threshold clustering of votes are successfully identified. After testing, the researchers found an effective recognition rate of 90% to 95% from the hand-held camera and a 60% to 70% rate from the humanoid robot. The researchers attribute this difference in performance to the lower resolution of the robot camera and thus a higher amount of noise. While the PCA-SIFT algorithm is generally successful in this experiment, the same results may not be obtained when the objects searched for are less complex, thus generating fewer features with which to identify the objects.

2.3 Usability

In addition to focusing on how to build a product, our team also completed tests to ensure our product is user-friendly. Lyons and Starner point out that before a product is complete, it needs to be tested for usability by its targeted consumers. Usability must be analyzed thoroughly via tests and surveys to determine if the system would benefit the community and also to differentiate the needs from the wants of the visually impaired. [11] There are many considerations in building such a device, including the camera mounting location, the comfort of the user, and the way the system looks and works. This article describes different usability tests our group could use to test our product, such as a capture vest (for capturing the video from user's point of view) and VizWear for analyzing user interaction. An experiment by Bradley and Dunlop [12] was designed to determine how the sighted and visually impaired think differently, especially when it comes to directions that can improve usability. The visually impaired have different spatial knowledge than the sighted population, so they need constant feedback about their surroundings and simple directions. Landmarks will be more helpful than just simple directions. [13] Although they may not be able to recognize landmarks such as a coffee shop, visually impaired individuals can locate crosswalks and lamp posts, to aid them in reaffirming where they are and where to move next.

Bane, Kolsch, Hollerer, and Turk [14] analyzed possible multimodial interactions with wearable augmented reality systems. The goal of their study was to provide roaming workers with advanced visualization equipment to improve situational awareness, ease, and effectiveness in their jobs. While wearable computers have evolved into tremendously versatile devices, traditional interfaces can only be as big as a device's surface. The authors proposed that the conflicting goals of device size and interface area could be met by expanding the interaction area beyond the device dimensions. For example, a head-mounted display could allow for information visualization in the entire field of view, extending far beyond its physical size. Additionally, they found that hand gestures performed in free space, recognized with a head-mounted camera, are not constrained to the hardware unit's dimensions. Combining these modalities could result in a more complex user-system interaction than is possible with a keypad. Certain features of the researchers' system, such as hand and voice recognition and object recognition, may be utilized in our product.

In addition to discovering what users might want in a system, it is equally important to find out what obstacles exist for the user. Marston and Church [15] identify specific barriers the visually impaired may encounter in navigation. They specify five types of spatial knowledge and environmental cues that the blind lack, including self-orientation and directional cues to distant locations. These barriers especially hinder the blind in new environments. These barriers must be considered when developing our navigational aid. Marston [16] studied the use of auditory cues in navigation by comparing navigation between individuals who were given auditory cues and those who were not. Marston deduced the blind are navigationally challenged due to a lack of spatial information. We want our product to overcome these obstacles and give the visually impaired user the knowledge he or she needs.

2.4 Existing Systems

The team explored currently available navigational systems in order to gain an in-depth understanding of the essential components of a navigational device and to better understand how to test a device's effectiveness. Analysis of these devices allowed us to develop guidelines for designing an improved navigational system. Each system that we explored exhibited fundamental flaws, such as cumbersome computer systems and inaccurate guidance aids. Much of the literature on navigational aids focuses on these flaws, allowing us to assess various approaches to designing an effective product and prepare for any problems we may encounter.

An article [6] that is a particularly useful guide for creating and testing a usable interface was written by David A. Ross and Bruce B. Blasch. In this study, researchers designed navigational aids for blind people, one for indoor navigation and one for crossing a street. The study's purpose was to identify the best types of user interfaces for a navigational aid system. Although, much has been done in this area of research, the authors write that little has been done to optimize user interfaces for an older population. To solve this problem, Ross and Blasch evaluated how effective interfaces were by measuring the number and type of user errors and the speed of the device. These were compared with results from baseline tests. The authors determined that a combination tapping-speech interface would work best for the majority of users, while a combination 3D beacon-speech interface may work best for a specific group. Though this study did not test computer-vision systems, it did study specific components being considered for our final design. The main drawback of this experiment was that testing was done in an urban environment, where the dynamics of street-crossing are somewhat different than they are on a college campus. A final limitation of this study was that all of the subjects were more than 60 years old. Our average user is significantly younger.

One of the ways our system would be most helpful to users is if the ability for user input is maximized. Krause, Smailagic, and Siewiorek [17] examined different ways for a system to adapt to its user's preferences. Integrating hardware/software design and user feedback is important so the user can utilize the product easily and accurately follow its directions. Novel software was developed using a two-fold input-output system integrating sensor data with user inputs to categorize and learn user preferences. The data was collected through a series of three experiments. The first measured the motivation of the machine learning approach via a user survey and threshold analysis. Studies two and three demonstrated the feasibility of the Context Identification method and preference learner using a self-report study and movement identification. A machine-learning method was found to be more suited for practical application. The researchers used a large amount of hardware, more than desired for our product, and the delay times were still around 10 seconds.

A study which used GPS as part of their system was authored by Loomis, Golledge, and Klatzky [18], who investigated the use of audio interfaces in a navigation system for the blind as a viable solution for communication between the consumer and program. The article describes the different approaches the authors used to incorporate the GPS system and an audio interface. Many technological advances, such as the improved GPS accuracy and virtual acoustic displays, could make a GPS system suitable for a navigational system for the blind. In this study, blind test subjects were each given a backpack mounted computer with GPS capabilities and a set of headphones with a motion sensing device attached to follow head movements. Each candidate was timed and observed going through different courses. Researchers analyzed the effectiveness of a GPS system and acoustic virtual display, comparing the time and accuracy of a subject's performance with what would be expected from a sighted individual. This research gave us valuable insights about how to integrate GPS and auditory user interfaces.

In addition to the use of audio signals to convey course correction to the system's user, it may be possible to provide much of this information via tactile feedback. A 2007 *Wired* magazine article entitled "Mixed Feelings" [19] discusses a system built by Udo Wächter, professor at the University of Osnabrück in Germany, that consisted of thirteen vibrating pads lining the inside of a belt. The belt was wired to a controller that was reading data from magnetoresistive sensors which were excited by the earth's magnetic field. The controller used these readings to vibrate the belt-pad which was closest to heading of north. Wächter wore this belt for six weeks and became accustomed to navigating with this additional sense and says that he "suddenly realized that my perception had shifted. I had some kind of internal map of the city in my head. I could always find my way home. Eventually, I felt I couldn't get lost, even in a completely new place." [19] This haptic sense he gained through the use of the belt could potentially be very useful to the users of our system. It may be possible to not only provide a device which would always orient the user to magnetic north, but could provide custom course headings based on their GPS route. This may enable the user to rely less on audio-based feedback, making it easier to hear their surroundings.

A study done by Liarokapis [20] combines a GPS device with computer vision techniques to create an augmented reality interface. The computer vision aspect detects predefined features on a route at City University in London. A 3-D model of the campus was built and used to present a personal view of the user's location (instead of a typical map overhead view) and to direct the user by superpositioning directions onto the camera image. An alternate method of using edge detection and template matching was also explored and future research includes integrating the two detection types. Because the scope of the project was limited to a specific area, matching was not processor intensive and was successfully implemented on a PDA. This research suggests that a combination GPS and computer vision system would be feasible as a blind navigation tool if a proper interface can be designed for it.

Chapter 3

Developing the Prototype

3.1 Interviews

During the preliminary coding and research of the three components in our visually impaired navigation system (CBIR, GPS, Audio), we interviewed members of the visually impaired community in order to understand if there was any initial interest in such a product. We conducted interviews on the University of Maryland, College Park campus with students of the Adaptive Technology Lab as well as with additional volunteers from the Columbia Lighthouse for the Blind on Route 1 in Riverdale, MD. The interviews were conducted by two members of our team, and all interviews were tape-recorded. In addition, a transcript for each interview was written up afterwards. When explaining our research study to volunteers, we answered questions for them before they signed off to take part in the study. These questions addressed topics such as an identification of the project title, the purpose and the procedures of our study, the assurance of confidentiality, a detailed description of the risks and benefits, the freedom to withdraw from the study, and the ability to ask questions.

Seven individuals (Five female and two male) were interviewed with ages ranging from twenty-four to sixty three and with visual impairments ranging from minor to severe. Some of our subjects were born with visual impairment and others acquired visual impairments over time. Causes of visual impairment included: a gunshot wound to the head, retinopathy of prematurity (ROP), retinitis pigmatosis, congenital/juvenile cataracts, glaucoma, diabetes, and sporting injuries. We asked the same thirteen questions to each and additional questions to gain further insight as needed. Our questions are included in Appendix A. This aspect of our study was crucial in moving forward to creating a practical and helpful device that will hopefully best suit the needs of the visually impaired community.

After conducting these interviews we analyzed our data to discern patterns in topics pertaining to usability of our system. We have learned that most prevalent current navigational tools are canes and people (mainly in unfamiliar situations and locations), and all subjects have shown high interest in our project. They also have stressed the importance of having options to navigate more independently. In addition, most subjects prefer a system that can locate destinations to one that locates specific lost objects. Those with limited interaction with sighted people focus more on functionality while those with high interaction with sighted people care about others' perceptions. However, our subjects have emphasized that for a system to be accepted, it must be unobtrusive and highly functional.

3.2 Hardware

3.2.1 Early Test System for Indoor Object Detection

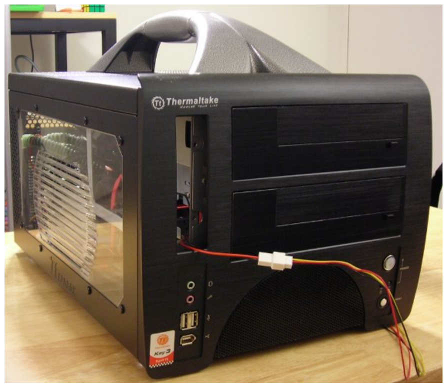
In initial tests of the indoor object identification and localization system, we utilized a Toshiba Satellite A75-S2311 notebook computer running Windows XP

Professional (32-bit) with Service Pack 2. The system had a 3.3 GHz Pentium 4 processor, 1.5 GB of RAM, a first-generation Logitech QuickCam for Notebooks Pro (mounted on sunglasses), a USB microphone, and over-the-ear headphones. While adequate for small-scale tests, that system was slow, bulky, energy inefficient, and plagued by tangling cables. Pentium 4 is an eight-year-old technology. Not only are modern multi-core processors better suited to handle our programs, but they are also far more energy efficient. With these thoughts in mind, we sought to build the most powerful system that the current commercially available hardware would allow. This would enable us to optimize our system to current technology and maximize algorithm speed and efficiency.

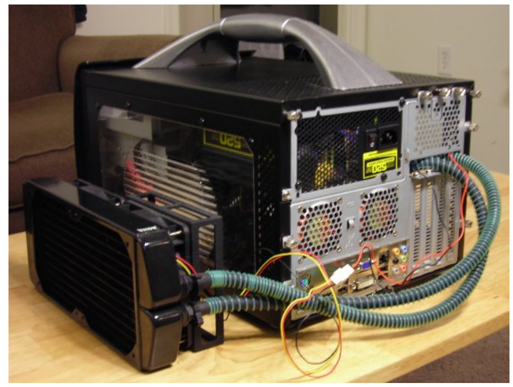
3.2.2 Benchmarking Desktop System

The computer

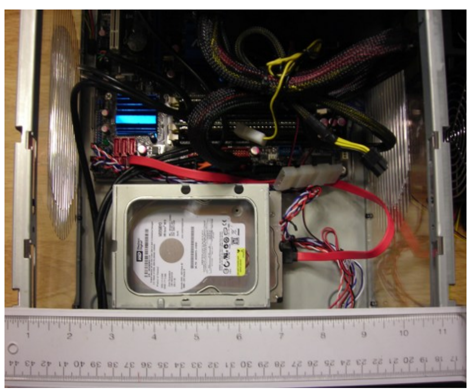
Given the computing demands of our computer vision algorithms, especially when combined with the added load of GPS calculations and constant audio interface, we realized a need for a powerful and advanced computer system to test and benchmark our applications. To this end, we custom-built a desktop computer (Figure 3.1) to maximize the processing power and see how fast we could run our software given the strongest available hardware. We chose a desktop over a laptop for this purpose because desktops offer significantly more powerful processing capabilities. Rather than purchase a retail computer, we custom built our system by assembling various components best suited for our purposes. This system utilized:



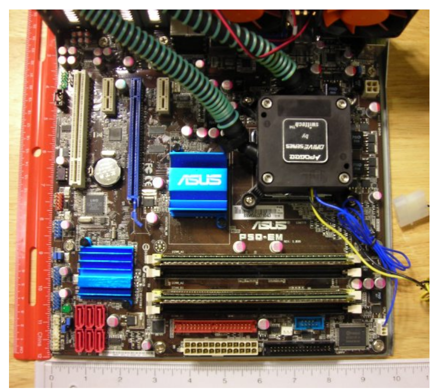
(a) Front of the case.



(b) Back of the case.



(c) Looking inside with the top cover off.



(d) Just the motherboard.

Figure 3.1: Benchmarking desktop system for computer vision applications, housed in a LANBOX gaming cube case.

- Intel Core 2 Extreme QX9770 3.2GHz Quad-Core Processor
- ASUS P5Q-EM LGA 775 Intel G45 HDMI Micro ATX Intel Motherboard
- OCZ Platinum 4GB (2 x 2GB) 240-Pin DDR2 SDRAM DDR2 1066 (PC2 8500)
- Patriot Memory PE32GS25SSDR 2.5-inch Internal Solid State Disk (32GB)

The processor, then the fastest available processor on the market, is a quad-core Intel CPU clocked at 3.2 Ghz. Besides being extremely fast, having four cores for parallel processing gives this CPU a significant speed advantage over the old system, which was a single-core Pentium 4. Processing power in the CPU is the rate-limiting step in all of our algorithms and so this CPU provides the greatest boost to our program's speed.

Initial plans were to custom build this system into a mobile computer for the blind, and to that end, we chose a Micro-ATX form factor motherboard because of its smaller, 9.6" x 9.6" size. While providing all the functionality needed in a mobile computer, this advanced Micro-ATX board nevertheless allowed us to retain all the advantages of a fast northbridge, southbridge, and front side bus (FSB).

Finally, we enhanced our hardware with four gigabytes of RAM to make sure that memory considerations were taken care of when handling large amounts of image and video data.

Because initial plans called for this system to be mobile, we selected a 32GB internal solid state disk (SSD) for our hard drive. SSD's, a relatively new development, are much like the flash-memory sticks that are now ubiquitous, allowing for large amounts of data storage in a light, compact, and robust unit. Given the physical stresses that the computer would take while walking around during testing, the increased hardiness of the SSD to physical shock is desirable. This is especially true when compared with the traditional Hard Disk Drive (HDD), which operates on a system of plates, ball bearings, actuators, and reader heads (much like a traditional record player), and is subject to disk failures given violent physical shock.

Difficulties

Difficulties arose in the building of this computer which eventually led to its adoption as a benchmarking desktop system rather than a true mobile computer. The primary issue was heat dissipation. With the 136 watt processor, the system generated a significant amount of heat, which, like in any computer, needs to be dissipated in order to function properly. Most commercial desktop computers are cooled by a fan-system, which unfortunately sits very tall on the motherboard and thus makes the overall computer large and bulky. We thus turned to a liquid-cooled option, which cools the processor by pumping cool water through an attached cooling block. This allowed us to reduce the height of the machine significantly, but also increased its overall bulk.

The second issue was power. Desktop computers normally run on an AC-DC power supply that connects directly to wall outlets with large current ratings. These power supplies are large and heavy, accounting for much of the weight of a standard computer. In a mobile environment, the primary power source would need to be a battery pack, similar to those carried onboard laptops. A DC-DC power supply, although substantially smaller than an AC-DC power supply would still need to be used, and would then connect to the motherboard through a 24-

pin ATX power connector. With the power demands of our machine, we needed a power supply with a rating of at least 250 watts, if not more, to ensure system stability and prevent crashes. This type of DC-DC supply is available, but the batteries required to operate the system would be impractically large. With an assumed power demand of 200W, a car battery would only be able to power the system for around three hours. This weight could be substantially reduced using high energy lithium cells but it seemed more prudent to wait for better battery technology or a more efficient processor before attempting to navigate using this computer.

In the face of these challenges, we decided to take this system, fully functional, if not very mobile, and dedicate it as a desktop benchmarking system to test the maximum speed of our algorithms and help develop software to use multi-core processors. For this function it has served very well, and we have achieved our fastest processing times on this computer.

3.2.3 A Navigation and Object Localization Device for the Blind

The computer

Our final system has to be light, mobile, compact, and user-friendly for a visually impaired person to use. Given these considerations, we decided to use a commercially purchased but custom configured notebook computer. For this purpose, we purchased a top-of-the-line Lenovo R400 Thinkpad (Figure 3.2). The Thinkpad was chosen for its no-frills design, light weight, and reputation for system reliability, hardiness, and long battery life. This laptop computer still retains significant computing power with a very fast processor and an abundance of RAM, but is small and light enough to wear in a backpack for mobile use. This computer served as the main testing and data collection for our experiments, and forms the core of our currently developed navigation device for the blind. The computer has the following specifications:

- Intel Core 2 Duo processor T9600 (2.8GHz 1066MHz 6MBL2)
- Intel Graphics Media Accelerator 4500MHD w/ 1394
- 4 GB PC3-8500 DDR3 SDRAM 1067MHz SODIMM Memory
- 160 GB Hard Disk Drive, 7200rpm
- Integrated Bluetooth PAN
- 9 cell Li-Ion Battery

While not quite as powerful as the QX9770 processor in the desktop system, this mobile T9600 still offers 2.8GHz of power in a dual-core system, allowing for optimum run speed for our algorithms, developed specifically to utilize multiple cores. We have the same 4 GB of memory, as well as a 160 GB HDD, allowing for the adequate storage of our databases of pictures, videos, and programs. Integrated Bluetooth was also added which allows us to use wireless Bluetooth connections for device peripherals, eliminating the need for clunky and distracting wires. Finally, a 9-cell battery was added, providing extended battery life to our system and enabling it to run for many hours on a single charge, a very helpful thing for any mobile device.

We have retained various features that come standard with any notebook computer despite their not being of high utility to blind users. For example, a

Figure 3.2: Lenovo R400 notebook computer.

production system designed for blind users, would not have a screen, CD-ROM drive, or any other such features which provide no added utility but result in increased bulk and weight. However, we have kept these features in our current system to make it easier to work with and test. Thus, the resultant weight of our system where only the most vital components are kept would be significantly lower.

The peripherals

In addition to the computer itself, various peripherals have to be connected and interfaced for our complete system to function.

For GPS localization, a Garmin 10x Bluetooth enabled module transmits data wirelessly to the system's laptop (Figure 3.3(a)). This lightweight, compact GPS unit allows the system to receive a stream of GPS coordinates and accuracy data, updating the user's location within the constructed map of campus. This GPS unit is ideal for use in the project due to its small size and weight (60g), wireless operation, and long-lasting lithium battery that charges though USB. In addition, this Garmin GPS uses the SiRFstarIII, one of the best commercial GPS chips available, with 20-channel architecture and support for WAAS and all satellite and ground based GPS transmitters. [21]

Additionally, we are using a TRX sentinel inertial navigation unit (INU), which functions as a dead-reckoning unit (Figure 3.3(b)). The INU contains various sensors that enable us to more accurately detect the user's position and heading. These sensors include three-axis solid-state magnetoresistors that are excited by the earth's magnetic field and enable us to determine the actual direction the user is facing, with respect to magnetic north. This is essential because heading determination using strictly interpolation of GPS coordinates is highly sensitive to drift. This would render a GPS-heading useless or potentially dangerous when the user is stationary or walking at low speeds. Other sensors include three-axis accelerometers as well as roll and yaw solid-state gyroscopes. We use these additional inputs to increase the accuracy of the calculated heading, compensating for tilt of the INU device that may occur during use. The inertial sensors will also allow us to extend the system in the future to stabilize the GPS readings and compensate for drift. Use of the INU, in conjunction with the GPS provides precise heading and location data.

To augment the audio-based interface, we built a tactile belt with an integrated microprocessor to relay heading information though vibration. The belt microcontroller consists of a Amtel AVR based microprocessor on an Arduino Duemilanove I/O board. Compass heading is read with a customizable baud rate from a standard Honeywell two-axis magnetic compass module with an on-board processor. Software we programmed on the MIPS processor determines the difference between the desired heading and the current heading and maps this angle to one of twelve vibration motors located around the perimeter of the belt to inform the user of the approximate angle they must turn to face in the specified direction. As the user begins to rotate toward the motor that is currently vibrating, the belt will give them continual heading feedback, allowing them to follow the precise direction required to arrive at their specified destination or an intermediate waypoint. The belt's electronics are enclosed in a clear plastic case around the size of a deck of cards, and the belt itself is designed to be covered by the user's clothing and as unobtrusive as possible. The most noticeable part of the belt is faint vibration sound near the same frequency and amplitude as a cell phone set to vibrate. A USB cable runs between the back of the belt and a hole in the backpack, providing power and serial communications between the belt's processor and the system computer.

- For object detection and localization, a Logitech quickcam for notebooks pro (Figure 3.3(c)), mounted to a pair of sunglasses (Figure 3.3(d)). A small camera allows for live video capture without being excessively heavy or bulky. Because of the large amount of data that must be sent through the camera, it must necessarily be wired.

- For user interface, a pair of Cyber Acoustics supra-aural headphones (Figure 3.3(e)), with an attached microphone for user voice-input. This combination headset-microphone allows the user to both send and receive information to the system using an intuitive speech-based interface.

- For user interface, a Nintendo Wii Remote (Figure 3.3(f)), providing a wireless push to talk button.

The complete system

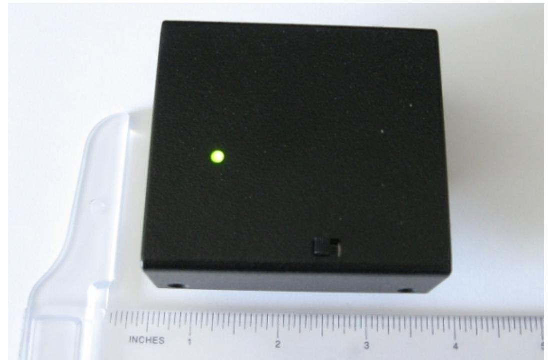
All the components connected together comprise our navigation and object localization device for the blind (Figure 3.4). The current prototype is comprised of the R400 notebook in a backpack, connected to its various peripherals. The GPS is clipped to the shoulder strap of the backpack to place it in the highest position. The INU is worn on the front of the body on a belt. Both of these devices interface wirelessly through Bluetooth. The camera is securely mounted to a large pair of black sunglasses, and connected to the computer in the backpack. The headphones are worn around the ears, and similarly connected to the computer. Finally, the Wii Remote can either be held separately, or attached to a common white cane, enabling the user to interface with the system with only one hand.

3.3 Interface

Our system's interface was designed with two primary goals in mind: First, our users would need an intuitive method of providing input and receiving computer generated responses; second, they would need a method of receiving directional guidance to locate destinations and other objects.

We decided that the best way to achieve the first goal was to use speech recognition and text-to-speech feedback. Our users communicate using speech in nearly all other aspects of their lives, so it seems natural that they should

(a) GPS Receiver

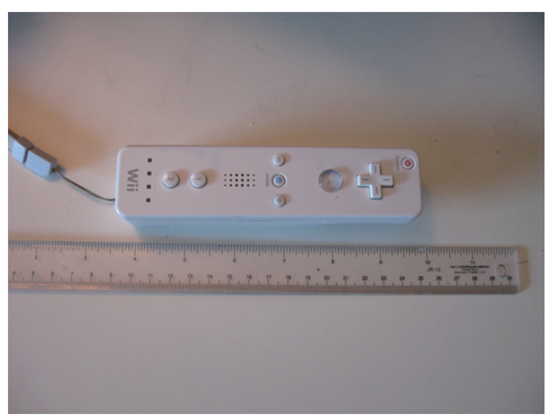


(b) Inertial Navigation Unit

(c) Webcam

(d) Mounted camera

(e) Headset



(f) Push-to-talk Button

Figure 3.3: Peripherals used with the prototype system.

Figure 3.4: Prototype build of our navigation and object localization device for the blind

be able to communicate with our system in the same way. Many robust speech recognition and text-to-speech algorithms have already been developed, and so our task was simply to use these algorithms to create an intuitive menu and response system. To this end, we determined which commands we would need to recognize, such as "go to the Student Union" or "find my cane". We then developed variations of these phrases to allow each user to speak as they would normally. The recognition system can decipher "go to the Student Union", "take us to the Stamp Student Union", "travel to Union", as well as several other versions of the same request. Natural language processing is a complex subject, but because the number of commands we need to recognize is fairly limited, we were able to include all of the most common variations.

We have taken several steps to improve the accuracy of our recognition process. Our system's limited vocabulary greatly improves its accuracy. Because it only needs to recognize a small subset of the English language, the recognition engine can identify commands much more accurately, even without the need for training that more complex speech recognition systems require. We are currently using the recognition engine built into Microsoft Windows applications such as Microsoft Office. This allows users to improve our system's accuracy even further by training the computer to recognize their voice.

To help prevent false positives by the recognition engine, we have experimented with several hardware options. We are currently using a push-to-talk button, an external button that is attached to the microphone or to the user's cane which can be held down while the user is issuing a command. The system will only recognize commands while the button is pressed, keeping the system from recognizing commands in ordinary speech and reducing the strain on the processor. We have also tried alternative microphones, such as the throat microphones commonly used in environments where the background noise is a problem. These microphones are placed in contact with the neck and pick up vibrations directly from the user's vocal chords. We intended to use a throat microphone to reduce the effect of background noise and improve recognition, but, while the audio quality produced by the microphones is acceptable for human use, we found that it was inadequate for computer processing. The degradation in recognition quality using current throat microphones outweighs the benefits of removing background noise, so, for the time being, we will be using a standard headset microphone.

To achieve our second goal for the interface system, that of guiding our users to a destination location or to a target object, we developed a few different feedback modes. The majority of these modes use audio to provide directional information, but the belt uses vibration to guide a user to his or her target.

Our first feedback mode has been used in other systems, and, while it is not the quickest method to use, it is intuitive and simple. We call this mode our "discrete feedback mode." We use the text-to-speech feedback previously mentioned to provide verbal cues that guide users to their destinations. For example, the phrase "right 30" might be used to tell a user to turn right thirty degrees. These directions can be updated once per second, or at any other interval specified by the user. The primary problem with using this method is that our users, lacking the precision of the computer guiding them, are unable to correctly gauge the angles given to them. They tend to overshoot the target and locate the target through trial-and-error. Users iteratively approach the zero degree line. The other problem with this method is that it requires a significant portion of users' attention to follow, distracting them from what is going on in the world around them.

Our second and third feedback modes are related, and can both be used in different situations. These are our "horizontal" and "vertical continuous feedback modes." Both modes use continuous audio beacons to guide a user to their target. The difference is that the horizontal mode is intended to guide users in one dimension, only focusing on left and right and ignoring up and down. This mode is especially useful in outdoor navigation, where the vertical directions are of very little of interest to our system. A user walking in an outdoor environment would require more feedback from the horizontal directions. The vertical mode provides feedback in both directions and is good for locating misplaced objects lying on the ground or on a table.

Both modes make use of three dimensional audio to provide directional information to the user. Humans are naturally able to pick out sounds from the world around them and identify the direction they come from quickly and intuitively. While we are not yet able to reproduce this precisely using current headphone technology, we can provide a reasonable facsimile using only the two channels provided by stereo headphones. The quality of this reproduction will depend greatly on the quality of the headphone, but even the cheapest stereo headphone can differentiate left and right. Well known algorithms exist for manipulating a sound in three dimensions so that when played back in stereo it sounds as if it was coming from the desired direction. These algorithms are commonly used for gaming or simulation, but they work just as well for our purposes.

These algorithms work best with a constant white noise, so we position a white noise sound in three dimensions around the user's head in the direction of their target. The algorithms work well to the sides of the user's head, but toward the front it can become difficult to differentiate the angle being represented. To this end, we provide a couple of different audio cues to guide the user to the centerline. The volume of the white noise is increased smoothly as the user rotates toward the desired direction. In the horizontal mode, a tone is also played that increases in both volume and frequency as the user approaches the center. Finally, a distinct popping sound is played whenever the user crosses the centerline, making it easy to identify the desired direction. In the vertical mode, we use the frequency to identify vertical direction, with a lower frequency used to identify the centerline and an increase in frequency indicating that the user is moving away from the centerline. This does not provide the same level of precision as the horizontal cues, but in most situations the vertical direction is much less important than the horizontal. The algorithm used to update the sounds used by the interface appears below:

```
function MoveSounds

float multiplier = 90.0 / StaticBounds
float horizDegrees = Abs(angleX)
float horizRads = horizDegrees * multiplier * (PI / 180.0))
int horizSign = Sign(angleX)
float staticX = horizSign * Sin(horizRads)
float staticY = horizSign * Cos(horizRads)
float vertDegrees = Abs(angleY)
float vertRads = vertDegrees * multiplier * (PI / 180.0)
int vertSign = Sign(angleY)
float freqY = vertSign * Cos(vertRads)
float freqZ = vertSign * Sin(vertRads)

if horizDegrees < StaticBounds then
      staticB3D.Position = new Vector3(staticX, staticY, 0)
      staticSound.Volume = 1.0 - horizDegrees
                              / StaticBounds * 0.3
else
```

```
        staticB3D.Position = new Vector3(horizSign, 0, 0)
        staticSound.Volume = 0.7
end

if horizDegrees > 0.5 then clicked = false
if angleX == 0 || (horizSign == -Sign(prevAngleX)
                        && horizDegrees < 45)
        if not clicked
                clickSound.Play(0, BufferPlayFlags.Default)
                clicked = true
        end
end

if FrequencyMode == "Horizontal"
        toneB3D.Position = new Vector3(staticX, staticY, 0)
        if horizDegrees <= ToneBounds
                toneSound.Frequency = lowFreq +
                        (highFreq - lowFreq) * (1.0f - horizDegrees
                        / ToneBounds))
                toneSound.Volume = 0.2 + 0.8 * (1.0 - horizDegrees
                        / ToneBounds)
        else
                toneSound.Frequency = lowFreq
                toneSound.Volume = 0.2
        end
else
        toneB3D.Position = new Vector3(0, freqY, freqZ)
        toneSound.Frequency = lowFreq + (highFreq - lowFreq) *
                        (1.0 - (StaticBounds - vertDegrees)
                        / (2.0 * StaticBounds))
        toneSound.Volume = 0.7 + 0.3 * (1.0f - vertDegrees
                        / StaticBounds);
end
```

A large part of the development of our continuous modes came about through trial and error. We tried different methods of providing directional feedback, initially with a repeating beacon sound in the desired direction and then with just the three dimensional white noise. We added the frequency and popping sounds as it became obvious that we needed more precision in the middle ten to twenty degrees. We continuously adjusted the volume, pitch, and angle parameters of our sounds until we arrived at a system that satisfied us. We then proceeded to integrate the audio system with vision and then navigation so that we could test how well it worked in actual applications and so that we could receive feedback from our users. See the integration section below for more information on this process, or the testing sections for more information on the experiments we used to test the interface in conjunction with the navigation and vision components.

Our final feedback method came about late in our development process, so it did not receive as much detailed testing or refinement as our audio modes, but from our initial tests and from the responses of those who have tried it out, it appears to be a promising alternative to our audio modes for directional feedback. The belt uses several vibrator motors positioned around the user to provide feedback and guide them in the correct direction. It has a greater concentration of motors around the front centerline to provide better precision while heading approximately in the desired direction. The belt is a good alternative to the audio feedback modes because it is less distracting and can be followed intuitively while still listening and responding to events in the surrounding environment. It can be used in conjunction with speech feedback to provide more complex directions and responses, giving the user better access to the information they need to function autonomously.

3.4 Navigation

3.4.1 GPS

For our system to provide navigation assistance, it first needs a start point, a destination point, and some way to construct a path between the two. We use data from a GPS to help users navigate along a path, traveling most of the way to their destination until they are close enough for other methods to provide terminal navigation. Our basic approach is to first obtain and parse GPS data, then feed it into our spatial representation of campus, and finally output a route that the user can follow to their destination.

Our system receives data over the bluetooth serial port from our GPS receiver. This data stream consists of sentences encoded using the NMEA 0183 standard. The most important sentences in this stream are those beginning with $GPRMC, sentences used to encode the user's latitude, longitude, and speed, among other pieces of information. This data allows us to locate the user relative to our map of campus. Once the user's current location is obtained, a route to their destination can then be calculated. Other NMEA sentences of interest include those beginning with $GPGSA and $GPGSV, which respectively include the precision information and the number of satellites in view. These sentences can be used to verify that the GPS is receiving data from satellites and to calculate the accuracy of the information being received.

We calculate a route between points using our spatial representation of campus. We use an API called Quickgraph, which provides us with the structures and functions needed to calculate the shortest route. A graph is a collection of nodes and the edges connecting them. The nodes are used to encode position information, while the edges contain a cost representing the distance between two nodes. We can take arbitrary points on campus and add them to our graph as nodes. We can then take traversable paths from node to node and add them to our graph as edges. In this way, our graph is a spatial representation of our campus. In practice, we have some points that are available as destinations (the Stamp Student Union, the Kim Engineering building, etc.). We also have many points that are used for intermediary routing but cannot be chosen as destinations.

We then use Dijkstras algorithm [22] to calculate the shortest path between our start and end points. The algorithm will return to us a route list with all the intermediary points on the optimal route. This information is then used to guide the user to their destination point by point. Dijkstras algorithm has also been proven to scale well as the size of the graph increases, so it will work well for future applications of our system.

Once a path has been found, the relative bearing between the user's current location and the next node in their path can be calculated. This angle, θ, can be calculated from the start and end latitudes and longitudes as shown in Equation 3.1.

$$d_{\text{Lon}} = e_{\text{Lon}} - s_{\text{Lon}}$$

$$\theta = \tan^{-1} \frac{\sin\left[d_{\text{Lon}} \cdot \cos(e_{\text{Lat}})\right]}{\cos\left[s_{\text{Lat}} \cdot \sin(e_{\text{Lat}}) - \sin(s_{\text{Lat}}) \cdot \cos(e_{\text{Lat}}) \cdot \cos(d_{\text{Lon}})\right]} \tag{3.1}$$

This system can be extended further by using other NMEA sentences that provide more information. In particular some sentences provide dilution of precision (DOP) information, which is a measure of accuracy of the GPS data. By using the DOP information, we can determine whether we can continue using the GPS for navigation or whether we need to transfer control to some other navigation system. We can also integrate this information with the data received from the INU, which would lead to a more accurate and reliable navigation tool.

3.4.2 Magnetic Compass

The INU used in our system contains a three axis magnetic sensor, which we use to calculate the user's current heading. Before a heading can be calculated, the system must first be calibrated by recording the minimum and maximum magnetic readings from each axis while rotating the INU in all directions. Using this calibration data, along with the current readings of the magnetic field in each of the two major axes (x and y), we can obtain the user's current heading using Equation 3.2.

$$\begin{aligned}
\text{Offset}:& \\
\delta_x &= \frac{x_{\max} + x_{\min}}{2} \\
\delta_y &= \frac{y_{\max} + y_{\min}}{2} \\
\text{Sensitivity}:& \\
E_x &= \frac{1}{\frac{x_{\max} - x_{\min}}{2}} \\
E_y &= \frac{1}{\frac{y_{\max} - y_{\min}}{2}} \\
m_x &= (x - \delta_x) \cdot E_x \\
m_y &= (y - \delta_y) \cdot E_y \\
\theta = \tan^{-1}&\left(\frac{m_y}{m_x}\right) + 180 \qquad (3.2) \\
\theta &= 360 - \theta
\end{aligned}$$

We then add or subtract 360 as necessary to obtain an angle between 0 and 360 degrees. This calculation will provide the user's current heading, with values of 0, 90, 180, and 270 representing north, east, south, and west, respectively.

Unfortunately, because this calculation only represents two axes of the magnetic sensor, small variations in the INU's tilt can have a devestating effect on the accuracy of the compass heading.

We can compensate for the error caused by the tilt of the INU by calculating and pitch and roll angles and using them to correct the compass heading. We use force of gravity on the INU's three accelerometers to calculate the pitch and roll angles in Equations 3.3 and 3.4 respectively.

$$\angle_{\text{pitch}} = -\sin^{-1}\left(\frac{a_y}{\sqrt{a_x^2 + a_y^2 + a_z^2}}\right) \tag{3.3}$$

$$\angle_{\text{roll}} = \cos^{-1}\left(\frac{a_x}{\sqrt{a_x^2 + a_y^2 + a_z^2}}\right) - 180 \tag{3.4}$$

We can then calculate the correct heading angle using Equation 3.5.

$$\delta_x = \frac{x_{\text{max}} + x_{\text{min}}}{2}$$

$$\delta_y = \frac{y_{\text{max}} + y_{\text{min}}}{2}$$

$$\delta_z = \frac{z_{\text{max}} + z_{\text{min}}}{2}$$

$$E_x = \frac{1}{\frac{x_{\text{max}} - x_{\text{min}}}{2}}$$

$$E_y = \frac{1}{\frac{y_{\text{max}} - y_{\text{min}}}{2}}$$

$$E_z = \frac{1}{\frac{z_{\text{max}} - z_{\text{min}}}{2}}$$

$$m_x = (x - \delta_x) \cdot E_x$$

$$m_y = (y - \delta_y) \cdot E_y$$

$$m_z = (z - \delta_z) \cdot E_z$$

$$x_h = m_x \cos(\angle_{\text{pitch}}) - m_y \sin(\angle_{\text{pitch}}) \sin(\angle_{\text{roll}}) - m_z \sin(\angle_{\text{pitch}}) \cos(\angle_{\text{roll}})$$

$$
\begin{aligned}
y_h &= m_y \cos(\angle_{\text{roll}}) + m_z \sin(\angle_{\text{roll}}) \\
\theta &= \tan^{-1}\left(\frac{y_h}{x_h}\right) + 180 \\
\theta &= 360 - \text{angle}
\end{aligned}
\tag{3.5}
$$

Once we have obtained the user's compass heading, we can subtract it from the bearing calculated as in the previous section to obtain a relative bearing angle for the user to follow. This angle is then passed along to the user using one of the modes discussed in the interface section.

3.4.3 INU

Acceleration data from the INU used in conjunction with the GPS would significantly increase the position estimates of the system. The GPS provides satellite positioning with an accuracy of about nine meters under good conditions. However, the accuracy decreases as the user moves towards a building and ceases to work as the user moves inside. In contrast, the INU is a dead-reckoning device and relies only on the previous position to determine the current position. INUs measure the acceleration of an object and estimating position is a simple matter of intergration. This has an advantage over the GPS because it does not require line of sight connections with satellites and can be used in any environment. Additionally, the resolution of the INU is much finer and is much more accurate than the GPS in short intervals. The downside is that dead-reckoning devices are useless without an initial position and are subject to drifts over time. Because previous data is used to calculate future data, errors are magnified as the process is repeated.

3.5 Computer Vision

3.5.1 SIFT Algorithm Description

The scale-invariant feature transform, developed by David Lowe in 1999, is a method for extracting information useful for identifying objects in an image, regardless of the size or orientation of the objects. This information, called *keypoints* or *features*, encapsulates the most noticeable geometric properties of an object, namely its corners and edges. Keypoints give SIFT its independence from variation in size, orientation, and illumination, and even partial object blockage in an image. Keypoints are detectable under a wide range of conditions. For example, keypoints can allow a door to be detected in an image taken on a cloudy day and is rotated 30 degrees clockwise.

A keypoint consists of four pieces of information: location in an image, scale, orientation, and descriptor. These data are determined from the input image, which is processed in four steps (Fig. 3.5). First, scale-space extrema detection identifies keypoints and defines their scales. Second, keypoint localization eliminates keypoints that are sensitive to noise, as well as keypoints that diverge from edges and corners. Third, orientation assignment defines a two-dimensional angle that is consistent between image rotations. Finally, the keypoint descriptor step defines a vector summarizing image gradients near a keypoint. The gradient of an image is the largest value of the discrete derivative at any pixel.

Figure 3.5: The complete scale-invariant feature transform process.

Scale-Space Extrema

Scale-space extrema detection has two sub-steps: construction of two image pyramids and a window operation on the images in the pyramid. In the first step of scale-space extrema detection, two image pyramids (Fig. 3.6) are derived from the original grayscale image (Fig. 3.8). The image spans the x-y plane. Copies of this image in varying sizes span the third dimension, called the scale-space (σ), at regular intervals. Each set of same-sized images along the scale-space axis is called an octave. Each octave consists of six images. The bottom-most image in the first octave, L_0, is the original image. As σ increases, images in the same octave become increasingly blurry, as when one magnifies an image beyond its original size. This blur effect is created by the Gaussian blur operator. The Gaussian blur operator is a function of the pixel values in the image and the location of the image on the scale-space axis, σ. As σ increases beyond the first octave, the next octave is reached. The first image in the next octave is constructed by taking every other pixel in the fourth image from the previous octave (L_3), halving the size of the image in both the x and y dimensions. The construction of octaves continues until the images are a maximum of 14 pixels in height or width (whichever occurs first). All of these octaves together are called the Gaussian pyramid. The difference-of-Gaussian (DOG) pyramid is computed by taking the absolute value of the pixel-by-pixel difference between adjacent images in the Gaussian pyramid.

After construction of the two image pyramids, a window operation on the images in the DOG pyramid is executed. Each pixel in a DOG image is compared with its 26 neighbors (Fig. 3.7). The 26 neighbor pixels are obtained by considering the three-by-three window around a pixel in D_i and extending it to

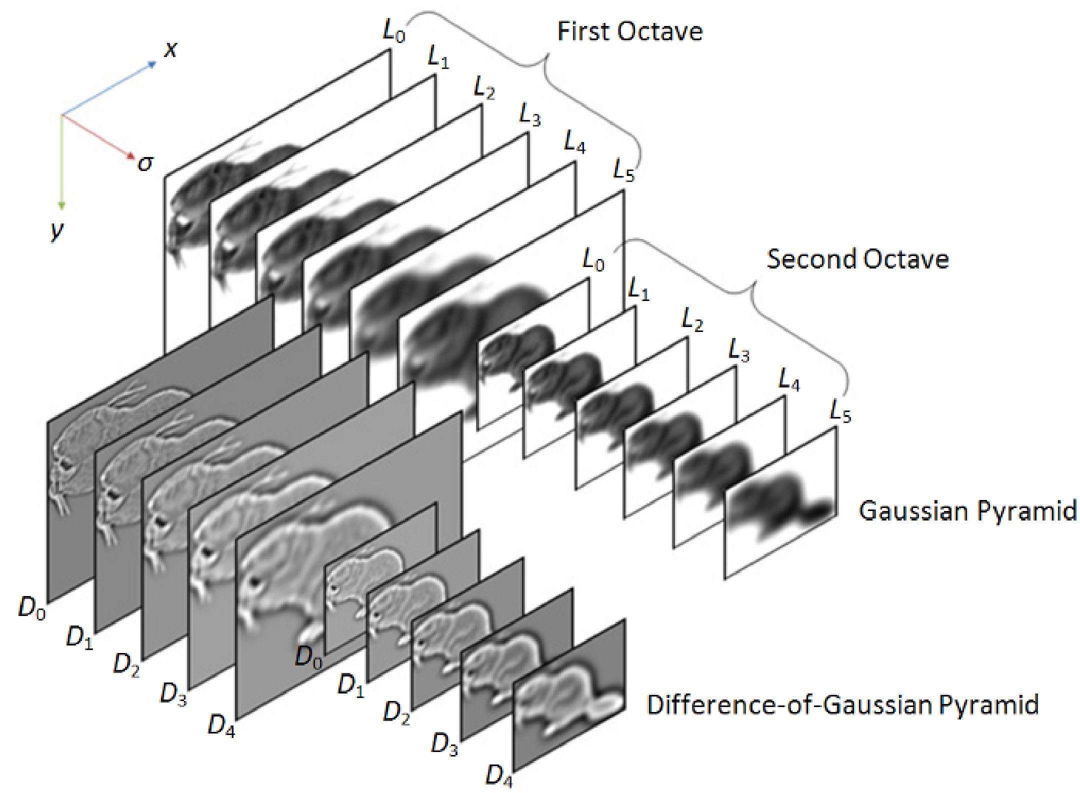


Figure 3.6: Two image pyramids required to extract keypoint candidates.

include $D_i - 1$ and $D_i + 1$, where $i = 1, \ldots, 3$. If the center pixel value is larger than its 26 neighbor values, then that pixel is a keypoint (Fig. 3.8). The scale, σ, at which the keypoint is found is assigned as the keypoint's scale.

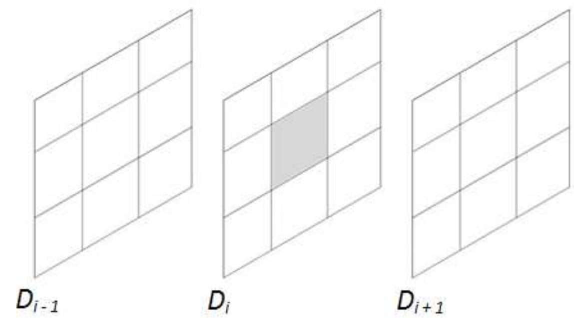


Figure 3.7: The 26 neighbor pixels (white) from the difference-of-Gaussian calculation.

In our implementation of scale-space extrema detection, a grayscale image $I(x, y)$ is used to construct the Gaussian pyramid. The i^{th} Gaussian-blurred image in the Gaussian pyramid, $L_i(x, y, \sigma)$, is given by a convolution of $I(x, y)$ with the Gaussian blur operator, $G(x, y, \sigma)$, where

$$L_i(x, y, \sigma) = G\left(x, y, k^i\sigma\right) * I(x, y), \tag{3.6}$$

and

$$G\left(x, y, k^i\sigma\right) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2k^{2i}\sigma^2}. \tag{3.7}$$

In the above equations, $k = 2^{1/3}$, $\sigma = 1.6$, and $i = 0, \ldots, 5$ for the six-image octave. Next, a DOG image pyramid is constructed by subtracting two adjacent Gaussian-blurred images. The j^{th} DOG image, $D_j(x, y, \sigma)$, is computed by

$$D_j(x, y, \sigma) = L_{j+1} - L_j, \tag{3.8}$$

where $j = 0, \ldots, 4$ for the five DOG images. Finally, for each pixel in $D_j(x, y, \sigma)$, the 26 neighboring pixel values are compared to determine if the pixel at (x, y) is a keypoint. In this window operation, image data within five pixels of the edge of an image are ignored. (Therefore, the minimum image area considered is 4×4 pixels.)

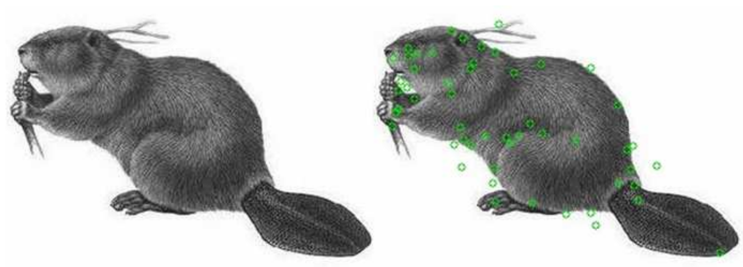


Figure 3.8: Keypoint candidates for the image shown to the left are shown as green circles to the right.

Keypoint Localization

After calculating the extrema or keypoints in an image of interest, the results are filtered based on contrast and location. To filter based on contrast, we use the equation

$$D(\hat{x}) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x}, \tag{3.9}$$

which calculates the offset between the sample point and the extremum. The threshold, $D(\hat{x})$ that determines whether or not a keypoint has enough contrast is set experimentally. Points that do not meet this threshold have low contrast and are sensitive to noise, so they are discarded from the set of keypoints.

To filter keypoints based on poor localization, such as those on an edge, we compute the curvature of the point and scale using the Hessian matrix

$$H = \begin{pmatrix} D_x x & D_x y \\ D_x y & D_y y \end{pmatrix}, \tag{3.10}$$

where the partials are estimated using neighboring sample points. Another threshold is determined for the ratio between the principle curvatures of the point, which can be used to determine whether or not the point is poorly localized on an edge. To check for a particular point, we use

$$\frac{(\alpha+\beta)^2}{\alpha\beta} < \frac{(r+1)^2}{r}, \tag{3.11}$$

with the appropriate value of threshold r and where α and β are the larger and smaller eigenvalues of H, respectively. Keypoints on an edge are sensitive to noise and are filtered out using this test.

Implementing these two tests increases the robustness of the SIFT features and results in better and faster matching. The two thresholds will be adjusted to balance how many features there are, how resistant the features are to noise, and how significant the features are.

Orientation Assignment

To achieve the rotation invariance that SIFT boasts, the rotation element of each feature needs to be removed. This is done using

$$m(x,y) = \sqrt{(L(x+1,y)-L(x-1,y))^2 + (L(x,y+1)-L(x,y-1))^2} \tag{3.12}$$

and

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right), \tag{3.13}$$

which need to be calculated on the Gaussian smoothed image, L, that matches the keypoints scale. Around each keypoint, a histogram of orientations weighted by magnitude and distance is made. The distance weight is calculated using a Gaussian distribution with $\sigma = 1.5$ times the scale. The largest peak in the histogram determines the direction of the feature point. In the case where there are multiple peaks (classified as being at least 80% of the largest peak), a feature is created at that point for each orientation. To improve estimation of the peak position, a parabolic curve is fitted to the keypoint and the two neighboring points and used to interpolate the final, dominant orientation. By aligning each keypoint along this angle, the resulting features become invariant to orientation.

Keypoint Descriptor

A keypoint descriptor is a 128-dimensional vector describing the gradients near a point in an image. The gradient information is rotated according to the angle found during the orientation assignment and weighted according to a Gaussian distribution with $\sigma = 1.5$ times the keypoint scale. The choice of the gradient is an emperically-derived conclusion based on human experiments. [23]

3.5.2 SIFT Feature Matching

To match SIFT features, we match their descriptors. As a first cut, to match a feature f to a collection of k features, f_k, we minimize the squared Euclidian distance between them, i.e., we find the feature in f_k for some particular k such that

$$\min \left[(f[0] - f_k[0])^2 + (f[1] - f_k[1])^2 + \ldots + (f[127] - f_k[127])^2 \right],$$

where the [ ] operator indexes the 128 elements of f and f_k. However, computing the squared Euclidian distance for each k^{th} feature in the set f_k requires too many operations for this algorithm to be useful. Instead, we use approximate methods. One approximate method to finding the minimum square Euclidian distance is the Best-Bin-First (BBF) algorithm. [23]

3.5.3 SIFT Reliability

The SIFT algorithm is designed to be robust to changes in size, orientation, and illumination. Because of the emphasis on using multiple keypoints for matching, the algorithm should also be robust to partial occlusion. However, this robustness is heavily dependent on the magnitude of the difference between the training images and the test images. Preliminary testing is used to set various thresholds and to design the most efficient device.

Several factors need to be considered to determine the optimal number of images needed in the SIFT database. Because the SIFT algorithm is processor intensive, populating the database with too many images would slow the system to a crawl. On the other hand, the more images there are in the database, the more features there are to compare with the test image and the more likely that a match will be found. During the testing process, different numbers of images in the database are used that encompass various sizes and perspectives. The optimum number is determined by the most number of templates the system can process in real-time and the difference between images is determined by the robustness of the SIFT algorithm. If SIFT can successfully detect an object from two times the distance the original image was taken at, the template images can be taken at those intervals. This method ensures that the overlap between templates is minimized and no processing time is wasted.

It is also important to reduce the number of false positives and false negatives generated by the SIFT operator. This is accomplished with template images containing enough features and with a properly calibrated matching threshold. Because SIFT features are highly distinctive, the probability of a mismatch between features themselves is low. As a result, false positives for an entire object is even less likely than for a single feature, unless the two objects look similar to begin with. Additionally, a larger number of features used to describe an object makes detection of that object more robust to changes in size, orientation, or occlusion, therefore reducing the rate of false negatives. Calibrating the threshold of matched features required for the detection of an object also increases the reliability of the system. By only requiring a certain percentage of the features in an image to be matched, the rate of false negatives can be reduced. While adjusting the threshold is typically a trade-off between the two types of error, SIFT's low chance of false positives makes this strategy especially effective.

In comparison with other local descriptors, SIFT boasts the highest accuracy, robustness, and descriptiveness. This makes SIFT a clear choice for feature based object detection in this system.

3.5.4 Other Features

A variety of other feature types and algorithms can be used depending on the situtation or application. For instance, color features attempt to identify objects by scanning an image for specfic wavelengths of light. Color features are a fast and efficient way of identifying objects with unique colors or patterns. The major weaknesses of these features are that objects do not always appear the same under different lighting conditions and that multiple objects might share the use of the same colors. Outdoor environments are especially unstable as the lighting and shadows from the sun vary throughout the day. On the other hand, a lit red exit sign has a unique color, is relatively consistent across various lighting conditions, and would be a good candidate for an object that can be detected using color features. Although, the objects typically carried by the blind may not be of a particular color, stickers may be placed on the objects to identify and differentiate between objects. The willingness of the blind to tag their possessions with these stickers is determined with user surveys.

Shape features are another type of feature that can be used to identify particular objects. The contours in an image that separate an object from its background can be used to identify the shape of an object. While the outlines are not necessarily discernable in the image, pre-existing knowledge of the shape can be used to facilitate the process. Using this method, objects with unique shapes such as humans or vehicles can be detected. Again, these features suffer from similar weaknesses as those in the color features. The outlines of objects may be extremely difficult to identify under different lighting conditions and objects with the same shape would incorrectly result as a match in the system.

These features have their own strengths and weaknesses and are not individually suitable for a blind navigation system. However, in combination with SIFT and PCA analysis, these detection algorithms can aid in the identification of various objects. Because the processing of color and shape features is much faster than SIFT, they can be used to pre-screen the image for a match and SIFT features or PCA can be used to confirm the match.

3.5.5 Data Collection

Subjects tested the integration of GPS, terminal navigation, and object detection in a navigational aid with audio feedback. They were asked to clip a 2.5 in x 3 in x 1 in compass and a 3 in. x 1.5 in. x 0.5 in. GPS to their waist (which together weigh half a pound) and a backpack containing a computer weighing 5 lbs. In addition, they wore a small microphone around their neck and a webcam mounted to their shoulder or to a pair of glasses. Subjects were asked to go from a designated point A to point B by telling the system where they want to go in an already marked path. The tests were conducted in a marked off area, away from vehicles, pedestrians, and anyone not associated with the project.

Before the test began, we trained subjects on how to use the system, taught them the voice feedback commands, and allowed them to get accustomed to the audio directions. There were pre- and post-surveys documenting their overall experience with the system.

3.6 Integration

The components of our system were developed almost completely separately from each other, with their final integration not occurring until the last months of the project. Each component was designed to function largely independently of the others so that it will be quite simple in future versions of our system to improve or replace the existing components or to add new components as they are created. Members of our group developed the computer vision system and audio interface system and integrated the global positioning and inertial navigation units. Each component was tested informally and reworked until the group members developing it were satisfied with its reliability and accuracy.

Once the interface had been developed, we entered the first stages of the integration process. Because the GPS and inertial navigation system was not yet ready to be tested, we began by combining a simplified portion of the computer vision system with the audio interface. Our goal was to test the effectiveness of the audio feedback along with the reliability of our computer vision implementation. The test was simple: we would ask the system to locate an object and once it was recognized by the computer vision algorithm we would provide the target object's relative angle to the user in real time using one of our audio feedback modes (Fig. 3.9). By trying each of the different modes and altering the settings within each mode, we were able to gauge our system's intuitiveness and accuracy and alter our approach accordingly. Initially we used only the horizontal continuous feedback mode, but we soon discovered that some vertical feedback was necessary for this type of test. If the target object was located on one of the vertical edges of the camera's field of view, then even if it was perfectly centered horizontally it would be difficult to track because it would move in and out of the frame. We repurposed the frequency sound to indicate vertical direction and noticed an immediate improvement in our results. It is likely that there exists an even better method of representing vertical direction using sound, and our system's modular design will make it easy to implement this new mode in the future.

Most of the difficulties we encountered during the integration of the vision system with the audio interface stemmed from the limitations of our testing hardware. The older technology we were forced to work with performed well enough for our purposes, but we were unable to run the vision algorithms in real time at the resolutions we had hoped to use. At lower resolutions, our target objects were still recognizable but because the computer had fewer features to work with, the types of objects recognizable by our system and the distances and lighting conditions we could recognize them from were more limited that they should have been. In conditions where the system was unable to consistently recognize a target object, our system became nearly unusable. The audio feedback would cut in and out as the system identified or lost the target object, making it next to impossible for users to locate the object. Once we received funding, we

Figure 3.9: Subject locating a white cane. In the upper right, a representation of the computer vision algorithm

purchased a newer, faster computer that eliminated most of the difficulties we had been having with the accuracy of our recognition algorithm. The new computer was also smaller, had better thermal properties, and had a longer battery life, making it much better suited to our needs.

The integration of the GPS and inertial navigation system with the audio interface proved to be of greater difficulty, not because the audio feedback was any more complicated but because of the underlying inaccuracies in both the GPS data and in data obtained from the compass and accelerometers in the inertial navigation unit. We began integration before the navigation system was complete, both because our time was growing short and because we needed a method of interacting with the computer while it was being carried in a backpack. Because the integration and the development of the navigation system happened simultaneously, we ran into problems that could have been avoided or at least made simpler if we had proceeded differently. For example, inaccuracies in the compass data due a slight tilt in the electronic sensors led to compass headings that were a bit off. We spent a great deal of time trying to fix portions of the code that dealt with integration, when the true solution was either to stabilize the compass so that tilt wasn't an issue or to correct the heading received from the compass using the inertial navigation unit's built in gyroscopes and accelerometers. We were able to develop a system that could guide users along the correct path to their destination, but the compounded errors from the GPS and compass led to a system that was much less accurate than we had originally hoped to achieve. Fortunately, it is possible to improve the stability of the system using data from the inertial navigation unit as well as from the computer vision system.

Because the inertial navigation unit we are using includes sensors for magnetic fields, accelerometers, and gyroscopes, we have all of the capabilities necessary to implement a system that improves the accuracy of the GPS data. This type of system that combines GPS data with inertial data to improve the accuracy and reliability of both is called a Kalman filter. Unfortunately we did not have enough time to implement this filter, but the framework exists in our system for this type of improvement to be added in the future.

The final stage of our integration was to combine all three systems into one. Our initial goal with this complete system was to implement a terminal navigation system, one that would use the GPS and compass data to guide the user most of the way to their destination and that would use the computer vision system to correct the user's direction once they approached their destination. Because the other two phases of integration were completed before we attempted this phase, achieving our goal was surprisingly simple.

The terminal navigation system works much like the GPS navigation system discussed previously, but once the user gets close enough to their destination the computer vision system is enabled. The distance at which this occurs varies depending on the location and the template images gathered for it, but for example in one of our tests the template images for a building were taken from about one, five, ten, twenty, and forty meters. In this case, the vision system would begin processing the feed from the camera once the GPS indicated the user was within forty meters of their destination. The vision system uses the same algorithm as in our object location experiment, processing the video feed to identify and locate the destination building. We use several template images from different distances and angles, and if necessary from different lighting conditions due to weather or time of day. Any direction received from the vision system overrides the direction provided by GPS and compass in order to guide the user to their target more accurately. If the target building is not detected, the system continues to function with the GPS and compass data. Once the user is within a couple of meters of their target, they are notified that they have reached their destination and the navigation and vision systems are disabled.

Chapter 4

Testing the Prototype

4.1 Component Test Results

4.1.1 Interface

Voice Recognition

The voice recognition system performed excellently. The speech interface allowed users to direct the computer in a number of different ways. For example, saying, "find cane," "find the cane," or "find my cane," would all result in the computer initiating a search for a cane. Even more variations can be added to make the feature as robust as possible. As the search is initiated, the computer gives confirmation, replying with "searching for the cane." If the computer misunderstands the user, a rare occurrence, the user simply presses the push-to-talk button and repeats the command. After the user has held the cane sign in the center of the frame for a configurable amount of time, the computer says "Cane found." Blind users liked that they could speak naturally to the computer, and that it would respond in plain English.

Directional Feedback

We evaluated two interface concepts. In one interface, a continuous sound is generated to guide the user to the object. We will refer to this as the 'beacon interface'. In a second interface, the computer outputs verbal instructions on how to angle the camera. Because the feedback from this interface is not continuous, we will refer to it as the 'discrete interface'. We tested both interfaces with a blind subject and obtained both quantitative and qualitative data. Quantitative data was obtained with two-dimensional sign location. Before testing, the subject familiarized himself with each interface by listening to the output as the computer tracked a target that he held in his hand (Fig. 4.1).

Figure 4.1: Subject training with the beacon directional feedback system.

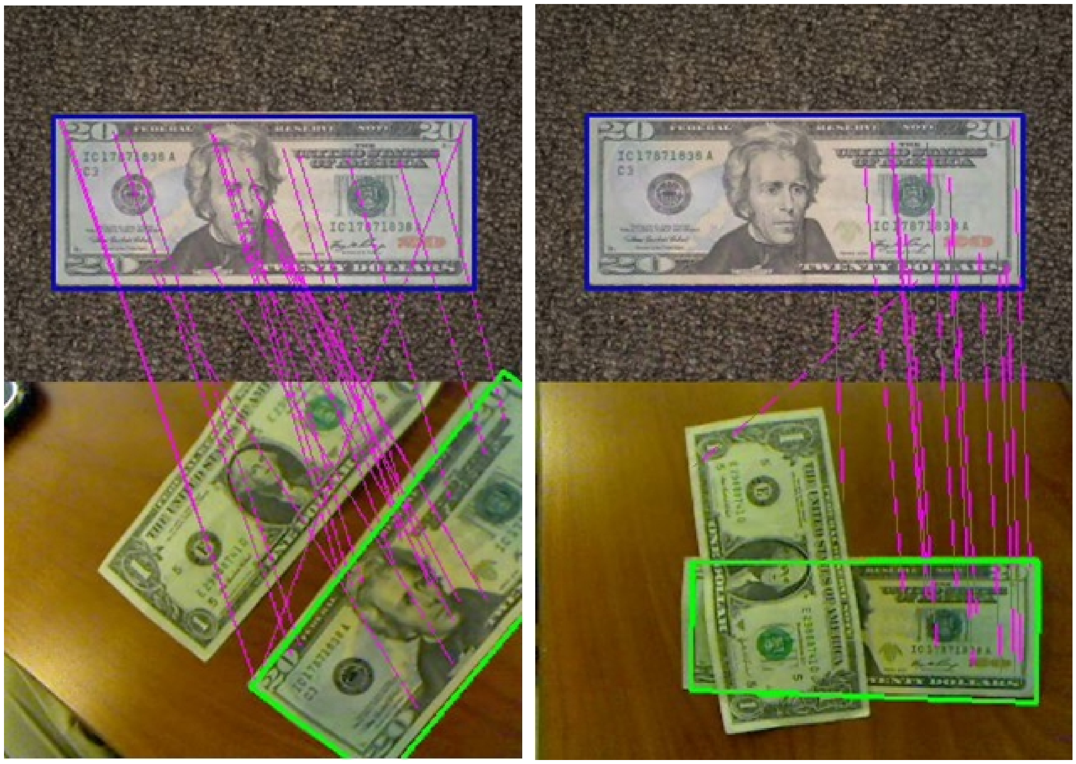
4.1.2 Reliability and Speed of Object Detection

In preliminary tests, which were done at near real-time rates, the algorithm successfully distinguished between a one dollar bill and a twenty dollar bill (Fig. 4.2(a)). In addition, the algorithm did not falsely identify the one dollar bill even when it occluded many of the most distinct features on the twenty dollar bill. In fact, it still successfully identified the twenty dollar bill (Fig. 4.2(b)). In the tests run so far, the algorithm has identified objects with exceptional accuracy.

However, there have been instances of the algorithm failing to find matches while an object of interest was in-frame. Even if the algorithm initially identifies the object, it does not always successfully track it when it or the camera is moved. Identification is not always continuous in real-time.

The camera is capable of providing the algorithm with thirty image frames-per-second; however, the algorithm easily consumes all of the processing power of our initial test system (Section 3.2.1) while only analyzing up to 20 frames per second.

Nonetheless, in three weeks of testing, there were no false positive matches. We tested two classes of objects. The algorithm had no trouble distinguishing between the signs in Fig. 4.3(a). In addition, during 'real-world' testing (Fig. 4.5), the algorithm never misidentified a cane, box of tissues, cup, or mug as any of the other objects.



(a) Rotated. (b) Occluded.

Figure 4.2: Even when rotated or partially occluded, the algorithm successfully distinguished between one and twenty dollar bills. Template images are shown above the video feed. The computer automatically generates the pink lines between corresponding features and maps the border of the bill with a green line.

4.2 Sign Experiment

4.2.1 Objective

Despite impressive research aimed at making signs more accessible to the blind [24, 25], infrastructure changes are expensive, and there will always be environments with limited accessibility. We are designing our system to be flexible enough to identify signs in such environments.

(a) Sign experiment setup. (b) Bird sign identification.

Figure 4.3: The subject was asked to point at the signs after the computer informed him that he had centered the image.

4.2.2 Setup

The subject was asked to center a sign on a wall in front of him both vertically and horizontally in the camera's view. After the subject held the image within three degrees of center for five frames, the computer would inform him that the sign had been found. He was then asked to point to the sign. The signs were then repositioned, and the exercise was repeated.

4.2.3 Results

SIFT Performance

We used three of the signs pictured in Fig. 4.3(a) throughout testing: the dog sign, the bird sign, and the restroom sign. The system had no trouble identifying the restroom sign continuously as the subject moved his head. The dog sign was identified with only a few discontinuities. In many cases, the bird sign could only be identified intermittently.

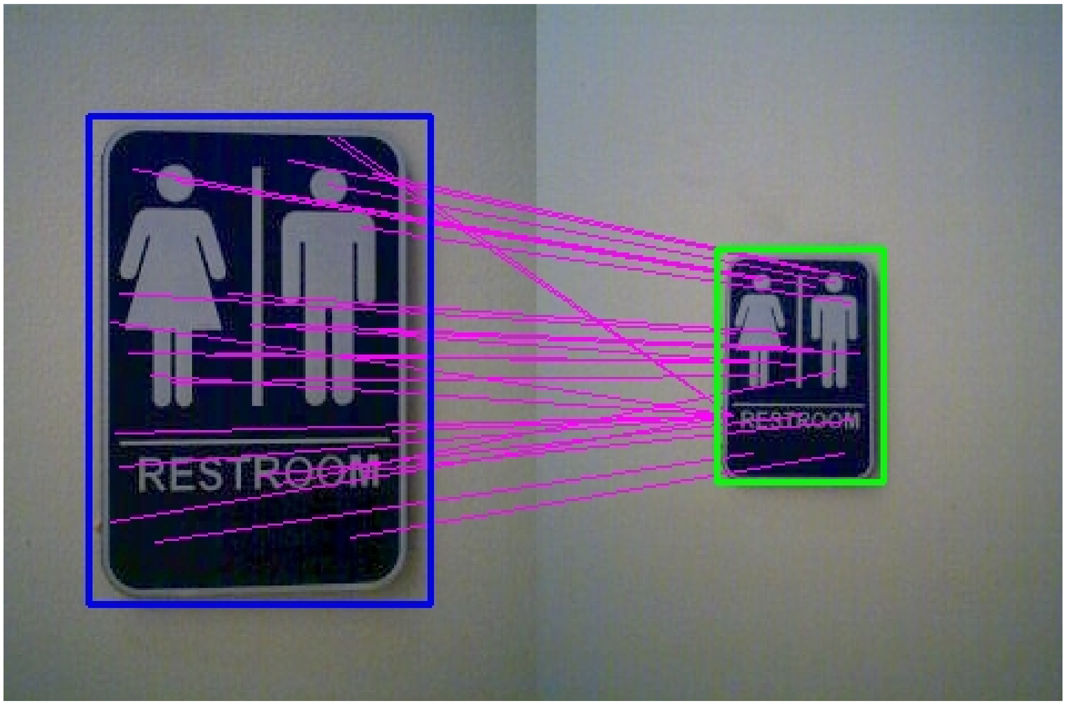


Figure 4.4: Feature matches from the template image (left) to the real-time video (right) of the restroom sign.

The algorithm is most successful when asked to identify signs with sharp, high-contrast features. It detects corners, but not curves. Both the restroom sign and the dog sign provide this type of feature. The gray background of the bird sign gives it a lower contrast ratio than the other two signs, and the curves of the bird's feathers are more difficult to detect than the edges of the restroom sign illustrations (Fig. 4.4) and the spots on the Dalmatian.

Localization with the Discrete Interface

Using the discrete interface, the subject was able to locate signs in an average of just over 16 seconds. This number includes any initial search time to get the signs in-frame as well as the time the subject had to keep the signs in the center of the frame. The signs were in-frame about 76% of the time.

When using the discrete interface, the subject tended to locate the signs iteratively. An instruction like "up 20 [degrees] right 4 [degrees]" would often prompt movement in the correct general direction, but of an incorrect distance. The subject would often overshoot the sign. In general, when using the discrete interface, the magnitude of the error vector would decrease with each iteration, but its direction would change dramatically.

Use of the discrete interface also led the subject to accelerate his head rapidly in the given direction, and then hold for instruction. The high jerk resulted in blurry frames and discontinuous identification. However, because the interface only needed information from one frame every time he moved his head, this did not impede localization.

Localization with the Continuous Interface

With the elimination of one outlier (see below), the continuous interface enabled sign localization in an average of under 12 seconds. Again, this number includes initial search time and object centering time. The signs were in-frame about 83% of the time.

When the algorithm could not locate a sign continuously, the interface was difficult to use, resulting in our outlier. The continuous beacon interface stops sending sound when the vision algorithm cannot locate the object of interest in three consecutive attempts (this number is configurable). It is designed to be continuous, and, when it is not, using it can be confusing.

However, when the algorithm recognized a sign continuously, the subject could locate it in an average time of just over 5 seconds using the continuous beacon interface. In addition, the subject exhibited a perfectly damped response. He did not overshoot the sign. Centering was accomplished in one smooth motion with little concentration or effort. As our vision algorithms improve, this interface should result in very rapid localization.

Subject Interface Preference

The subject preferred the beacon interface when object identification was continuous and the discrete interface when it was not. His preference stems from his heavy reliance on environmental sounds for information. Use of an overly distracting interface could impair his ability to keep himself safe. After little practice, he could use it intuitively and automatically. The discrete interface required more attention to use, and would interfere with his ability to process environmental sounds.

4.3 Object on Table Experiment

4.3.1 Objective

In addition to identifying environmental signs, blind users have indicated that they would like to be able to use our system to identify and locate misplaced objects. This experiment was designed to evaluate the vision algorithm's ability to locate 3D objects and the audio interface's ability to help a blind user find them.

4.3.2 Setup

After the sign tests, the subject sat at a table that had a uniform cloth draped down from the wall over its surface. Objects and people did not cast significant shadows in the environment, which was lit with overhead fluorescent fixtures. A modified white cane (Fig. 4.5), a cup (or sometimes mug), and a box of tissues were quietly placed in front of the subject. He was then asked to identify one of the objects and center it within three degrees of vertical and horizontal in the camera's view. After the object was centered for 5 frames, the computer would inform the subject that the object had been found. He then reached out and touched the object *without* searching with his hand.

Figure 4.5: Subject locating a white cane.

4.3.3 Results

The vision algorithm successfully identified three-dimensional objects, though under tightly controlled conditions. Successful identification is highly dependent upon the object templates. In order to account for all possible object orientations, a significant number of templates must be used. Unfortunately, increasing the number of templates too much can slow down the algorithm. Image resolution can also become a constraint. The hardware (Section 3.2.1) had difficulty processing images larger than 320×240 pixels in real-time. This limited the number of visible features on the object when it was not close enough to the camera.

When suitable objects were placed in the same orientation as the template images, within the limited focal depth of the camera, and close enough that a 320×240 frame could capture enough features, the algorithm performed well and identification was continuous.

4.4 Indoor Exit Location

4.4.1 Objective

Indoor navigation based solely on the computer vision algorithms is an integral aspect of the system that will aid the visually impaired in locating objects of interest within a room. Such important objects of interest are doorways, exit signs, chairs, and any of the user's personal belongings. Door and exit location is especially important in providing access points for the visually impaired to enter and exit various environments. Validating the ability of the system to locate and lead users to doorways was established as an important first step in testing the capabilities of the system. Initial experiments were conducted using sighted individuals as preliminary measurement of the success of the system.

4.4.2 Setup

In this experiment, the goal was to locate, travel to, and open a static door over twenty feet away within a room using only the system and the aid of a white cane. To ensure independence from test to test, the test subjects were blindfolded during each trial and disoriented at the beginning of each trial such that their relative orientation to the door was arbitrary. After initialization of the system, the progress of the test subject was tracked to measure the time needed to detect the door and the time needed to subsequently travel to the door and open it. Sixteen trials were obtained, using three trials from five testers and one trial from one other tester.

4.4.3 Results

Over all sixteen trials, the user was able to successfully detect, move to, and open the door using the system and a white cane. The average total time of each trial was 38.1s, representing a relatively short time needed to locate and move to an exit. The average time of door detection was 17.2s compared to an average time of 20.9s needed to then travel to and open the door. However, there was a considerably greater amount of variability in the time to detect the door, representing the randomness of the relative starting orientation of the test subject to the door. Indeed, the time to door detection had a range of up to 55.0s, while the time needed to travel to and open the door had a smaller range of 22.0s. Further, the challenge of traveling without sight contributes to the longer average time after door detection.

There is further variation in each individual test subjects average times after door detection. While some users were new to the system, others had been exposed to the audio interface and thus were more able to quickly travel to and open the door. The lower average times after door detection represent these relatively trained individuals, while the higher average times display those that were less familiar with the system. This suggests that the system will become more effective as the user becomes accustomed to the system and lead to quicker location of doors with repeated use.

4.5 GPS and INU Short-Range Outdoor Navigation

4.5.1 Objective

The combination of short-range, computer vision terminal navigation and long-range GPS navigation is the ultimate goal of our project. The system that we designed should be able to guide users both in a gross sense (from building to building) and in a fine sense (from an area in front of the building to the entrance door). To test our system for this application, we organized example routes to walk, mainly in front of the Jeong H. Kim Engineering Building (Kim building), and examined the accuracy of our system in guiding users to the door.

4.5.2 Setup

For short-distance navigation experiments, pre-selected routes were plotted in the digital map for walking to the door of the Kim Building from various starting points.

Gemstone Team Vision members themselves served as test subjects for these preliminary experiments (Figure 4.6, page 72). In order to try to replicate the experience of a fully blind user, test subjects wore a full mask blindfold. Subjects were also given a standard white cane to feel the area immediately in front of them for dangers and hazards, much as a blind person would do. The system (minus the camera and sunglasses) was worn in its entirety, as described in the hardware section (Figure **??**). Subjects stood at a preset location and indicated through the audio user interface that they wanted to go to the Kim building. Subjects then followed the audio cues, given through the previously described continuous beacon interface, to the destination. Fellow team Vision members accompanied the test subject to monitor their progress and ensure safety. These experiments served as proof-of-concept runs for our system.

4.5.3 Results

In preliminary tests, GPS accuracy varied from trial to trial, but was generally acceptable. The system acknowledged the reading of a map node within 1-4 meters of its actual location. The location of map nodes, even between trials, stayed relatively constant, indicating that the GPS was not fluctuating in its reading, but that the map nodes just had to be moved on the digital map. Sometimes, within our margin of error, the node would be located in the middle of the street or in the grass next to a sidewalk. Subjects reported that the audio interface was generally easy and intuitive to use, though it took some getting used to. In subsequent trials, as subjects became more accustomed to the interface, their ease of use was improved.

Figure 4.6: Blindfolded team Vision member Lee Stearns using the device to follow a pre-determined path to the front door of the Kim Engineering Building. Other team members are nearby to ensure his safety.

Navigation with the assistance of the INU was generally accurate, but was highly dependent on the pitch, roll, and relative orientation of the INU. Vertical tilt of the INU had a great effect on early tests, and initially, mechanical methods were sought to correct for this. Eventually, algorithmic compensations were created for this tilt by writing programs for tilt-correction. While the system generally led users in the correct direction, a field of error of anywhere from 5-15 degrees was observed, often resulting in users straying from the path and going into the grass or other areas. The white cane helped users avoid general obstacles when walking, as well as allowing users to follow the curb of a street to ensure that they stayed on the sidewalk. Thus, the results of early GPS-only experiments were promising, showing that the system functioned well in a general sense, but required some fine-tuning to be more accurate.

4.6 GPS and INU Long-Range Outdoor Navigation

4.6.1 Objective

Experimental procedures for the short distance outdoor navigation were repeated but applied to a significantly longer course. Here, our aims were to demonstrate proof-of-concept of our system as a campus navigation aid.

4.6.2 Setup

The user wore the same system as described above, including the computer system, GPS, INU, headset, and white cane (Figure 4.7). Test subjects (again team Vision members) walked a pre-plotted path from the Stamp Student Union to the Kim Building, or the other way around, calculated from our campus map and shortest-path-algorithm.

4.6.3 Results

Again, the system performed well for the gross navigation task. Nodes were reached within a few meters of their actual location. Navigation along long sidewalk routes was sometimes complicated by inaccurate directional headings, sometimes directing users slightly off course. However, we found that we could

Figure 4.7: Blindfolded team Vision member Lee Stearns using the device to follow a pre-determined path from the front door of the Kim Engineering Building to the Stamp Student Union. Other team members are nearby to ensure his safety.

compensate for this by having users rely on the white cane to feel for the edges and curbs of sidewalks, which any blind user can follow with ease. During testing, subjects would often switch to reliance on the white cane along long paths, and utilize the GPS guidance of our system upon reaching nodes requiring a change in direction, or in following paths that do not follow along the side of a road.

4.7 GPS and INU with Computer-Vision-assist Short-Range Outdoor Navigation

4.7.1 Objective

Initial outdoor tests used only the GPS and INU system to test gross navigation. We then integrated this GPS system with the Computer Vision system used for indoor Navigation, creating an outdoor navigation system that guided users both to the general area of the entrance of a building (with GPS) and then closer to the specific entrance door, using computer vision. Thus, our goals with this test were to navigate users as close to the door as possible, right up to the door handle, and do so using the combination of GPS and Computer Vision.

4.7.2 Setup

Experimental setup was the same as described above in Section 4.5, with the addition of the sunglasses and mounted webcam for use with computer vision applications. Thus, for these tests, the subjects were wearing our system in its entirety. The camera wire was looped behind the users head and into the computer in the backpack.

4.7.3 Results

With the addition of the camera, navigational accuracy was much improved, especially closer to the door of the building. In general, the GPS accuracy fades as you near large buildings, which interfere with its signal. However, this is also true when computer vision techniques become most powerful, as proximity to features increases. Thus, in these tests, as users neared the building entrance, the camera took over navigational guidance from the GPS and was able to provide highly accurate terminal guidance to the door. This was similar to results achieved in the indoor "door-finding" tasks described earlier. Overall, results from these tests were very promising and proved that our system could accomplish the tasks it was designed for.

Chapter 5

Discussion

5.1 Future Capabilities

The future capabilities of the system largely deal with scalability. One area that may not scale very well is the image processing. The image processing algorithms currently scan the entire picture database and try to feature match using SIFT, but this approach is currently used with a modest database size. The size of the database could very quickly increase if the system was to be used as a general purpose navigation system instead of one confined to the University of Maryland campus. This is due to the need for many templates for all possible test images that need to be processed. As previously noted, a large database size would slow the SIFT algorithm down considerably. Some optimizations that deal with other information available to the system are needed to solve this problem.

The proposed solution is to limit the amount of pictures in the database to only those needed by SIFT at any given time. This would be done by tagging the images with their GPS location and then using the current GPS location to only use the images that are in close proximity to the user. Intuitively, it makes sense to partition the database so that if you are near a particular building, the system only uses pictures that are from the area near that building. It would not be efficient to search though the entire database. Such an approach could lead to false-positive matches with far-away but similar looking buildings. With the GPS-tagging approach, the image database could grow to thousands of images, but SIFT would look at the only the images closest to the user in determining what the test images is.

In using this new approach, two things are needed: The current GPS position and GPS tagged images in the database. The GPS location is something that is already heavily used for navigation and would incur no additional cost to use. The tagged database can be constructed using a special camera or specialized hardware that automatically produces such images. There are already commercially available products that can perform this task. The increased information encoded in the images is small in comparison to the size of the images and should present no additional storage challenge beyond that of the images themselves.

The scalability of any computer vision system is always questionable due the intense processing that computer vision algorithms require. With future advances in computing hardware and the use of optimizations such as the one detailed above, we are confident that our system could scale to very large sizes. Furthermore, even if the system is not scaled in such a way, optimizations make it possible to do more complex processing on images and are useful for further research.

5.2 The Gemstone Program

Through the Gemstone program, we were able to bring together a team of undergraduate students from a variety of different majors in order to tackle a significant problem in our community. Due to the size and diversity of our team, it was initially a challenge to find the best roles for each member. Nevertheless, within the past three years, we were able to come together to create a useful and innovative navigation aid for the blind. Through experiences such as GEMS202 and Team Gemstone, we were better able to understand the strengths and weaknesses of our teammates, as well as figure out a way to solve problems as a team. Benchmarks that Gemstone program has set throughout the past three years, such as the junior colloquium, have allowed us to refine our presentation skills and kept us on track in terms of research.

Goals for the foreseeable future would require expertise from students from a variety of different backgrounds. One possible direction includes creating an interactive guide for local attractions. We are extremely close to the nations capital and its wealth of museums, and monuments. Future improvements to our system would allow the users to take a guided tour of whichever historical sites are programmed into the system. Not only could it efficiently guide users through the mazelike confines of a museum, it would also provide them with information regarding the exhibits they visit. Once outside, hungry users could ask the system could direct them to a local restaurant.

Each passing month brings about advancements in consumer electronics that could be used to significantly improve the performance of this navigation system. Many of the weaknesses of the current system can be addressed by combining the software with more advanced hardware components that are certain to come out in the near future. A lighter and more powerful laptop would improve the performance of the system while simultaneously decreasing the weight strain on the user. A wireless webcam that is integrated into a pair of glasses would improve the appearance of the system and eliminate unnecessary cables. A more sensitive GPS receiver would increase the reliability of guidance, while integrating a wireless push to talk button into the white cane could streamline the system and reduce the amount a blind user would have to carry. Finally, the future team could work on efficiently incorporating WiFi into our system, which would give it access to the wealth of information available on the internet.

5.3 Conclusion

Our approach to solving the problem of navigation for the blind differs greatly from companies making similar products. It is our belief that by working with the blind community at every step, we will be able to create a product that provides what they actually need in a navigation tool. By integrating computer vision and GPS tracking software in a cooperative manner, our system will provide users with greater knowledge of their surroundings. Designed with input from the blind community, our product will give users the information they need, when they need it. It is our hope that the development of this technology will enrich the lives of the visually impaired, giving them the independence to travel on foot to places they have never visited before. Coming to the U.M. College Park campus as a freshman can be an overwhelming and challenging experience, even for those who have the luxury of sight. By providing those without sight a means to travel independently on campus, we will be helping them take part in the University of Maryland experience.

Appendix A

Interview Questions

A.1 Initial Interviews

1. How old are you?
2. Were you visually impaired your whole life? If not, how long have you been visually impaired?
3. What caused your visual impairment?
4. What do you currently use to navigate?
5. Have you used any other navigational aids in the past? What were they?
6. Is it important to you to have a device that may help you identify and locate destinations? objects?
7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?
8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

9. How important is it to locate and track the movement of people?
10. What kind of items/objects do you need to be able to recognize from a distance? (Examples: doors, people, etc.)
11. What kind of personal items might you need to locate in case you drop or lose something? (Examples: keys, wallet, etc.)
12. How do you feel about marking personal belongings to make it easier for the camera to detect them? (Example: putting stickers onto cell phones or wallets.)
13. How do you feel about wearing cameras or other equipments on certain parts of your body? (Examples: Head, shoulder, arm, waist, etc.)

Appendix B

Interview Transcripts

B.1 Subject 101, Female

B.1.1 Interview Questions

1. How old are you? 24

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   No, 5 years.

3. What caused your visual impairment?

   The subject survived the gunshot wound to the head.

4. What do you currently use to navigate?

   Cane and people (walking with somebody)

5. Have you used any other navigational aids in the past? What were they?

   None other than people. She has been offered to have an eye dog, but the subject declined the offer because even though the eye dogs were well trained, she couldn't trust a dog to be outside navigating around. She prefers PEOPLE.

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   Yes. a device that helps me identify and locate destinations or objects would make me more independent.

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   She does not specify the items that she needs to find frequently. She said that she's been using her cane to find her stuff pretty easily. And since she always leaves her stuff in the same place, she has no problem finding and locating things. In other words, NO specific items.

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

   The subject states that it would take her some time to get used to it. This device would be just additional guide to her cane and people. Even with any devices, she says that she would still walk around the campus with her cane. She wants a device that could tell her where to turn, where the stairs, the streets, and curves are in addition to the directions.

9. How important is it to locate and track the movement of people?

   It's very important because she says that she walks around everywhere, and she doesn't want to bump into anybody especially since UMD is a huge campus with a lot of kids. She also points out that there would be no need of navigational device because people around her(note-she always walks around with someone right beside her.) can just help her direct or guide her if there's another person walking towards her.

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc. In addition to people and cars, doors would be important because she would have to know where to go to. Stairs, curves, and ramps would be important.

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.
    Mostly her personal items but whatever little things that she has trouble finding. Her technique to find her stuff is to have her cane lay flat on the ground and brush it over the ground side to side. Then the objects that she's looking for would be in contact with the cane.

12. How do you feel about marking personal belongings to make it easier for the camera to detect them? Example: putting stickers onto cell phones or wallets.
    It would be okay. Stickers are okay as long as they are not too big or intrusive.

13. How do you feel about wearing cameras or other equipments on certain parts of your body? Examples: Head, shoulder, arm, waist, etc.
    As long as they are not big, obvious, and noticeable, it's okay

B.1.2 Suggestions and Comments

Headsets would be helpful. Just one ear because the person needs to keep the other ear open to the environment. Portable device.

The device that doesn't get wet when it starts raining suddenly. (Water proof)

Not too noticeable

big laptop=inconvenient

She would definitely try to see how reliable our device would be.

B.2 Subject 102, Male

B.2.1 Interview Questions

1. How old are you?

   25

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   Yes, I was.

3. What caused your visual impairment?

   I was born with it.

4. What do you currently use to navigate?

   Just a cane. I use people for rides or if I'm with someone I'd grab there arm and they would lead me, which is a lot faster than a cane if I'm not familiar with a certain area.

5. Have you used any other navigational aids in the past? What were they?

   No.

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   Yes.

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   Cell phones, wallet, cane, keys, clothing.

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

   It would be very important for a person to locate a place because it would give a person a lot more freedom so that he wouldn't have to rely on someone to take him around if not familiar with a certain location. No preference whether the query is typed or spoken.

9. How important is it to locate and track the movement of people?

   It depends if I'm in a crowded area.

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc.

    Buildings, to know if I'm approaching a destination. It would be beneficial to recognize a pedestrian walkway.

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.

Cell phone, keys, cane, wallet, money (change)

12. Do you have a specific technique of finding things you misplace?

    Not really. I have sometimes used a cane to feel for things.

13. How do you feel about marking personal belongings to make it easier for the camera to detect them? Example: putting stickers onto cell phones or wallets.

    That would definitely be beneficial. I'd rather that whatever it is on the item be smaller, but color does of the marker doesn't matter.

14. How do you feel about wearing cameras or other equipments on certain parts of your body? Examples: Head, shoulder, arm, waist, etc.

    It depends on how noticeable/big something is. An ideal size of something to wear is the size of a clip on mic, and I would put it on a collar, a shirt pocket, or a pants pocket.

15. Would you ever consider not using your cane?

    If navigational aid was really good I would consider not using the cane. I go to places that I'm very familiar with without a cane (examples are a friends house, 9:30 club, Black Cat)

16. Do you prefer a male voice or female voice?

    No real preference. I'm used to a male voice, but I could get used to a female voice. No preference about pitch.

17. Is controlling the speed important?

    Yea, I would definitely like to have that option.

B.2.2 Suggestions and Comments

Headsets are a good idea. Just one ear so I can hear what's going on around me. Subject would like to come back and test the system

B.3 Subject 103, Female

B.3.1 Interview Questions

1. How old are you?

   55

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   Yes

3. What caused your visual impairment?

   Retinopathy of Pre-maturity

4. What do you currently use to navigate?

   Sense of touch (sun on my face), sense of hearing - also a hearing impairment, no current Navigational aids. Cane. People, but only in social situations and in complicated places.

5. Have you used any other navigational aids in the past? What were they?

   Dog guides

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   Yes - destinations. Yes - objects might not be a bad idea

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   cell phone

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

   Pretty important. Prefer verbal communication with system because it's easier while walking

9. How important is it to locate and track the movement of people?

   Sometimes important - if I ever have grandchildren; in a crowded room and you need to find a specific person;

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc.

    Doors, specific rooms in buildings, signs, bathrooms, water fountains

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.

    cell phone, keys, wallet

12. How do you feel about marking personal belongings to make it easier for the camera to detect them? Example: putting stickers onto cell phones or wallets.

    No problem whatsoever

13. How do you feel about wearing cameras or other equipments on certain parts

Doesn't bother me at all. Ideal size of wearable system = nothing bigger than Braille Note

B.3.2 Suggestions and Comments

None

B.4 Subject 104, Female: 11/1/07 2:53 PM

B.4.1 Interview Questions

1. How old are you?

   54

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   No, I had perfect vision until 10 years old and then lost a lot vision in past 10 years

3. What caused your visual impairment?

   RP (retinitis pigmatosis) and I took vitamin A and K in huge doses, but since Vitamin A is toxic in large amounts I was advised to get off it and since then I have lost a lot of my vision.

4. What do you currently use to navigate?

   Cane. Metro Access

5. Have you used any other navigational aids in the past? What were they?

   No

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   Yes. Yes

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   Keys, Cane, Buttons, Quarters. I get on my hands and knees if I can't find them

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

   That would be very worthwhile to ask the system to find something and I would prefer verbal communication because I have arthritis in my hands

9. How important is it to locate and track the movement of people?

   Very Important, that way I don't bump into people - usually I don't bump into people when using my cane. (would be open to stop using cane if the system was very good)

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc.

    Doors, Buildings

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.

    Keys, Coins

12. How do you feel about marking personal belongings to make it easier for

the camera to detect them? Example: putting stickers onto cell phones or wallets.

No Problem, nothing too big and nothing too small

13. How do you feel about wearing cameras or other equipments on certain parts of your body?

    Ideal size = size of cell phone (nothing too big); and be able to put it on hip

B.4.2 Suggestions and Comments

None

B.5 Subject 105, Female: 11/1/07 3:15 PM

B.5.1 Interview Questions

1. How old are you?

   63

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   Born with cataracts, visually impaired since birth

3. What caused your visual impairment?

   Congenital cataracts, followed by Glaucoma

4. What do you currently use to navigate?

   Long White Cane

5. Have you used any other navigational aids in the past? What were they?

   Science Products for the Blind's cane that makes noise depending on where it is perceiving other objects. The Sonic Torch

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   I would really welcome it so as not to have to rely on other sources such as people which are second best. Yes (to objects)

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   Tree branches (in the rain they get weighted down), pole, and other objects that are above the vicinity of the cane. In regards to finding objects - "I don't misplace them"

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

   That would be great. That would be marvelous. Prefers this to typing in a query

9. How important is it to locate and track the movement of people?

   There are occasional situations where it's hard to focus on one particular person. For instance, it would be nice to track people that you are having a conversation with while following them

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc.

    Buildings, Entrances, Traffic Crossings, Poles, Cars in a parking lot, trees

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.

    cane, purse, coat

12. How do you feel about marking personal belongings to make it easier for the camera to detect them? Example: putting stickers onto cell phones or wallets.

    Feel good about that, especially on luggage at the airport. Color preference or conspicuousness is not important

13. How do you feel about wearing cameras or other equipments on certain parts of your body?

    Major concern is the fragility of the object, therefore exposure to the elements and size is important only in that might be break if it is too large. Desktop Keyboard is too big. Ideal Size = walkman, or something that can easily be placed in a purse

B.5.2 Suggestions and Comments

One headphone is adequate

Loves the idea of the system conveying the information verbally as opposed to tones and rings

B.6 Subject 106, Female: 11/1/07 3:38 PM

B.6.1 Interview Questions

1. How old are you?

   25

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   Since 2001

3. What caused your visual impairment?

   Diabetes and hit in the eye with a football

4. What do you currently use to navigate?

   My mobility, my vision is adequate to navigate during the day, but I use other people to help me navigate at night

5. Have you used any other navigational aids in the past? What were they?

   No

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   Yes. Yes

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   Keys, glasses, many things around the house

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a

query?

That’s a great idea. It’s easier, quicker to verbally ask the system than typing it

9. How important is it to locate and track the movement of people?

   I can pretty much see them now, so a device is not really necessary

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc.

    Doors (especially if they are partially open), Poles, Signs (the low ones like “Wet Floor”)

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.

    Keys, Glasses, Earrings, Change,

12. How do you feel about marking personal belongings to make it easier for the camera to detect them? Example: putting stickers onto cell phones or wallets.

    It would be nicer if things came integrated with a sticker/sensor type of device already so as not to have to buy 200 stickers and place them all over objects. “It’s going to help me see. I don’t care what people say”

13. How do you feel about wearing cameras or other equipments on certain parts of your body?

    Similar to a necklace pendant, or a half dollar. Definitely something around the neck = no need to fidget with hands, pockets, purse, and able to be hands free. Nothing on the head

B.6.2 Suggestions and Comments

None

B.7 Subject 107, Male: 11/1/07 3:38 PM

B.7.1 Interview Questions

1. How old are you?

   61

2. Were you visually impaired your whole life? If not, how long have you been visually impaired?

   Yes. Officially declared blind by age 2

3. What caused your visual impairment?

   Juvenile Cataracts

4. What do you currently use to navigate?

   Cane, "pretty women," occasionally another person to help me

5. Have you used any other navigational aids in the past? What were they?

   A GPS system manufactured by Sindero. It was very useful for a few reasons: I knew exactly how far destinations were. It helped me navigate in unfamiliar areas, such as helping to find close restaurants.

6. Is it important to you to have a device that may help you identify and locate destinations? Objects?

   Yes. Yes

7. Are there specific items that you may need to find frequently (i.e. cell phones, cane, wallet, ID, clothing, etc.)?

   Cell phones, clothing with the appropriate colors I will need to match what I'm wearing

8. How important is it to have the ability to ask the system to find something or some place? If it is important, how would you most want to perform a query?

   Fairly important - It would be very useful to find specific room in buildings such as a bathroom, or a specific room number inside a larger building. Prefer typing queries into a system for a number of reasons: 1)because if it's too noise in a certain area it might be difficult to speak, 2) Don't want others to know my business, 3)Prefer to keep things private

9. How important is it to locate and track the movement of people?

   Not very important

10. What kind of items/objects do you need to be able to recognize from a distance? Examples: doors, people, etc.

    Signs, street signs, when the traffic light changes, where bus stops are. More interested in finding destinations than lost objects since I travel a lot

11. What kind of personal items might you need to locate in case you drop or lose something? Examples: keys, wallet, etc.

    Cell phones, wallet, check card. Finding clothing is unimportant because I'll just go to my clothing drawer, but finding the appropriate colored clothing to match what I'm wearing is important. I'm frustrated I don't have a lot of choices to express clothing options (he currently shops for ties by how they feel and the mood that this feeling gives of. He calls it his "aesthetic touch").

12. How do you feel about marking personal belongings to make it easier for the camera to detect them? Example: putting stickers onto cell phones or wallets.

    Doesn't feel comfortable calling attention to himself, interfering with distorting other people's vision, or anything that feels too conspicuous

13. How do you feel about wearing cameras or other equipments on certain parts of your body?

    He wants "something to enhance and not interfere", therefore an ideal size would be something the size of a pin, a cell phone (because you can put it in your pocket). Basically something that is utilitarian, but unobtrusive to outsiders because that would call too much attention to him.

B.7.2 Suggestions and Comments

Wants to be able to control volume since he fears that a loud, audible noise will bother others and he doesn't wish to be rude or disruptive. He feels that a positive outcome of this system is that it might finally allow him to give back and help others with directions and navigational information

Appendix C

Consent Forms

C.1 Consent Form for Interviews

Project Title	*Navigational System for the Visually Impaired*
Why is this research being done?	Our goal is to create a device that will improve the mobility of the visually impaired using a combination of GPS, mounted cameras, and image processing. We plan to produce a system that will vastly increase the navigational information available to its users. This information will be available through an interface that will be simple for visually impaired individuals to use. Our efforts will be focused on answering the following research question: how technology can best be used to address the unmet navigational needs of the visually impaired community on the University of Maryland, College Park campus?
Why is this research being done?	Our goal is to create a device that will improve the mobility of the visually impaired using a combination of GPS, mounted cameras, and image processing. We plan to produce a system that will vastly increase the navigational information available to its users. This information will be available through an interface that will be simple for visually impaired individuals to use. Our efforts will be focused on answering the following research question: how technology can best be used to address the unmet navigational needs of the visually impaired community on the University of Maryland, College Park campus?

What will I be asked to do?	Participants will either come to the University of Maryland campus to complete our tests, or we will go to them if they cannot find transportation. The procedure involves one 45-minute to one-hour session consisting of four separate parts. This involves a survey/questionnaire in which the participants' answers will be recorded on an audiocassette tape. The survey/questionnaire will take place in the offices of the University of Maryland, College Park or in a closed room if the survey/questionnaire is done off campus.
What about confidentiality?	We will do our best to keep your personal information confidential. To help protect your confidentiality, we will keep all files in a locked room and computer data entry on password-protected computers. Additionally, you will be assigned a number and will be referred to by that number for any and all data entry and data interpretation to be used only by gemstone mentors and members of the gemstone team. If we write a report or article about this research project, your identity will be protected to the maximum extent possible. This research project involves making audiotapes of you for the purposes of recording your answers to our questions. The tapes will be stored in our team's office and only gemstone mentors and members of the Gemstone Team will have access to the recorded tapes. Please sign if you agree to be audio taped during your participation in this study. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.
What are the risks of this research?	There are no known risks.
What are the benefits of this research?	While there are no immediate personal benefits, your participation will help us develop a navigational aid that best suits the needs of the visually impaired community.

Do I have to be in this research? **Can I stop participating at any time?**	Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.
What if I have questions?	Professor Rama Chellappa and Team Vision at the University of Maryland, College Park campus, are conducting this research. If you have any questions about the research study itself, please contact Roni Tessler at: The University of Maryland, 5300B South Campus Commons, College Park, MD, 20742 or at 301-802-1218 or `rtessler@umd.edu`. If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, Maryland, 20742; (e-mail) `irb@deans.umd.edu`; (telephone) 301-405-0678. This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.
Statement of Age of Subject and Consent	Your Signature indicates that: • you are at least 18 years of age; • the research has been explained to you; • your questions have been answered; and • you freely and voluntarily choose to participate in this research project.
Signature and Date	**NAME OF SUBJECT**
	SIGNATURE OF SUBJECT ___ I agree to be [videotaped / audiotaped / photographed] during my participation in this study. ___ I do not agree to be [videotaped / audiotaped / photographed] during my participation in this study.
	DATE

C.2 Consent Form for Integrated Interface and Computer Vision

The procedure will ask participants to qualitatively rate different forms of directional information in regards to the audio component of our system.

Project Title	*Navigational System for the Visually Impaired*
Why is this research being done?	Our goal is to create a device that will improve the mobility of the visually impaired using a combination of GPS, mounted cameras, and image processing. We plan to produce a system that will vastly increase the navigational information available to its users. This information will be available through an interface that will be simple for visually impaired individuals to use. Our efforts will be focused on answering the following research question: how technology can best be used to address the unmet navigational needs of the visually impaired community on the University of Maryland, College Park campus?
What will I be asked to do?	Participants will either come to the University of Maryland campus to complete our tests, or we will go to them if they cannot find transportation. The procedure will ask participants to qualitatively rate different forms of directional information in regards to the audio component of our system.

What about confidentiality?	We will do our best to keep your personal information confidential. To help protect your confidentiality, we will keep all files in a locked room and computer data entry on password-protected computers. Additionally, you will be assigned a number and will be referred to by that number for any and all data entry and data interpretation to be used only by gemstone mentors and members of the gemstone team. If we write a report or article about this research project, your identity will be protected to the maximum extent possible. This research project involves making audiotapes of you for the purposes of recording your answers to our questions. The tapes will be stored in our team's office and only gemstone mentors and members of the gemstone team will have access to the recorded tapes. Please state if you agree to be audio taped during your participation in this study. Or if you do not agree to be audio taped during your participation in this study. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.
What are the risks of this research?	There are no known risks in this study.
What are the benefits of this research?	While there are no immediate personal benefits, your participation will help us develop a navigational aid that best suits the needs of the visually impaired community.
Do I have to be in this research? **Can I stop participating at any time?**	Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.
Is any medical treatment available if I am injured?	The University of Maryland does not provide any medical, hospitalization or other insurance for participants in this research study, nor will the University of Maryland provide any medical treatment or compensation for any injury sustained as a result of participation in this research study, except as required by law.

What if I have questions?	Professor Rama Chellappa and Team Vision at the University of Maryland, College Park campus, are conducting this research. If you have any questions about the research study itself, please contact Roni Tessler at: The University of Maryland, 5300B South Campus Commons, College Park, MD, 20742 or at 301-802-1218 or `rtessler@umd.edu`. If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, Maryland, 20742; (e-mail) `irb@deans.umd.edu`; (telephone) 301-405-0678. This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.
Statement of Age of Subject and Consent	Your signature below indicates that: • you are at least 18 years of age; • the research has been explained to you; • your questions have been answered; and • you freely and voluntarily choose to participate in this research project.

C.3 Consent Forms for Integrated System

Subjects will test the integration of GPS, terminal navigation, and object detection in a navigational aid with audio feedback.

Project Title	*Navigational System for the Visually Impaired*
Why is this research being done?	Our goal is to create a device that will improve the mobility of the visually impaired using a combination of GPS, mounted cameras, and image processing. We plan to produce a system that will vastly increase the navigational information available to its users. This information will be available through an interface that will be simple for visually impaired individuals to use. Our efforts will be focused on answering the following research question: how technology can best be used to address the unmet navigational needs of the visually impaired community on the University of Maryland, College Park campus?
What will I be asked to do?	Subjects will test the integration of GPS, terminal navigation, and object detection in a navigational aid with audio feedback. They will be asked to clip a 2.5in $\times$ 3in $\times$ 1in compass and a 3in $\times$ 1.5in $\times$ 0.5in GPS to their waist (which together weigh half a pound) and a backpack containing a computer weighing 5 lbs. In addition, they will wear a small microphone around their neck and a webcam that will either be mounted to their shoulder or to a pair of glasses. Subjects will be asked to go from a designated point A to point B by telling the system where they want to go in an already marked path. The test will be conducted in a marked off area, away from vehicles, pedestrians, and anyone not associated with the project. Before the test begins we will train subjects on how to use the system, teach them the voice feedback commands, and allow them to get accustomed to the audio directions. Finally, there will be a pre- and post-survey documenting their overall experience with the system.

What about confidentiality?	We will do our best to keep your personal information confidential. To help protect your confidentiality, we will keep all files in a locked room and computer data entry on password-protected computers. Additionally, you will be assigned a number and will be referred to by that number for any and all data entry and data interpretation to be used only by gemstone mentors and members of the gemstone team. If we write a report or article about this research project, your identity will be protected to the maximum extent possible. This research project involves making videotapes of you for the purposes of recording your answers to our questions. The tapes will be stored in our team's office and only Gemstone mentors and members of the gemstone team will have access to the recorded tapes. Please state if you agree to be taped during your participation in this study. Or if you do not agree to be taped during your participation in this study. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.
What are the risks of this research?	Team Vision members will be monitoring all subjects during testing at all times to prevent any risks. The testing will be done in a marked off area to prevent other people from entering our testing area. There are no known risks.
What are the benefits of this research?	While there are no immediate personal benefits, your participation will help us develop a navigational aid that best suits the needs of the visually impaired community.
Do I have to be in this research? **Can I stop participating at any time?**	Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

Is any medical treatment available if I am injured?	The University of Maryland does not provide any medical, hospitalization or other insurance for participants in this research study, nor will the University of Maryland provide any medical treatment or compensation for any injury sustained as a result of participation in this research study, except as required by law.
What if I have questions?	Professor Rama Chellappa and Team Vision at the University of Maryland, College Park campus, are conducting this research. If you have any questions about the research study itself, please contact Roni Tessler at: The University of Maryland, 5300B South Campus Commons, College Park, MD, 20742 or at 301-802-1218 or `rtessler@umd.edu`. If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, Maryland, 20742; (e-mail) `irb@deans.umd.edu`; (telephone) 301-405-0678. This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.
Statement of Age of Subject and Consent	Your signature below indicates that: • you are at least 18 years of age; • the research has been explained to you; • your questions have been answered; and • you freely and voluntarily choose to participate in this research project.

Appendix D

Quantitative Data

D.1 Indoor Exit Location

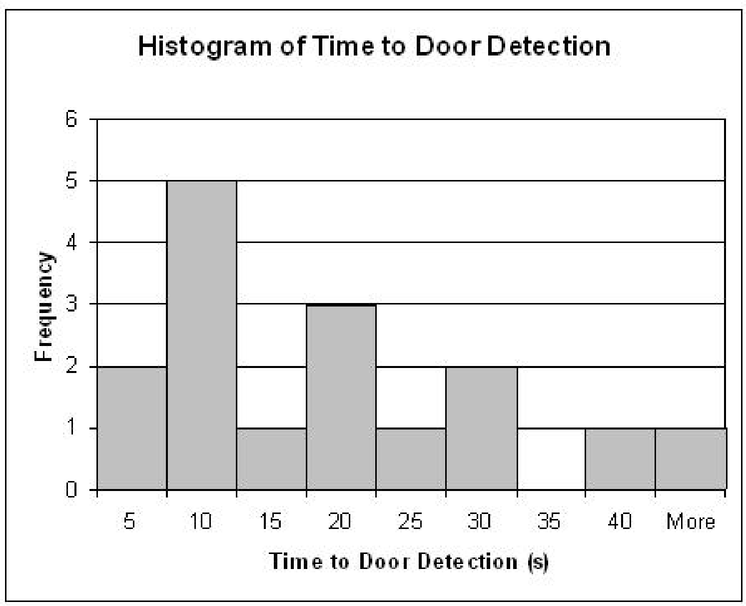


Figure D.1: Distribution of times to door detection: A large amount of variation was observed in these times, representing the randomized relative starting orientation during each trial.

Subject	**Total t (s)**	**t to Door Detection (s)**	**t after Door Detection (s)**	**Ave. t after Detection (s)**
1	48.80	25.80	23.00	23.00
2	43.00	11.00	32.00	27.33
2	27.36	6.36	21.00	
2	44.38	15.38	29.00	
3	87.00	57.00	30.00	24.03
3	51.70	38.00	13.70	
3	34.90	6.50	28.40	
4	49.00	21.00	28.00	24.00
4	31.10	3.10	28.00	
4	25.30	9.30	16.00	
5	22.91	6.91	16.00	12.33
5	13.00	2.00	11.00	
5	37.44	27.44	10.00	
6	25.35	9.80	15.55	16.02
6	35.26	17.76	17.50	
6	32.68	17.68	15.00	

Table D.1: Test results for locating, traveling to, and opening a static doorway using the system: Data was collected for time needed to identify where the door was from the starting location and time needed to then travel to the door and open it. With some variation, the tests demonstrated the ability to successfully locate the door in real-time. Variation lay in the arbitrary relative starting orientation with respect to the door that is reflected in the time to door detection as well as familiarity with the system. As users became more accustomed to the system, decreases in time after door detection were observed.

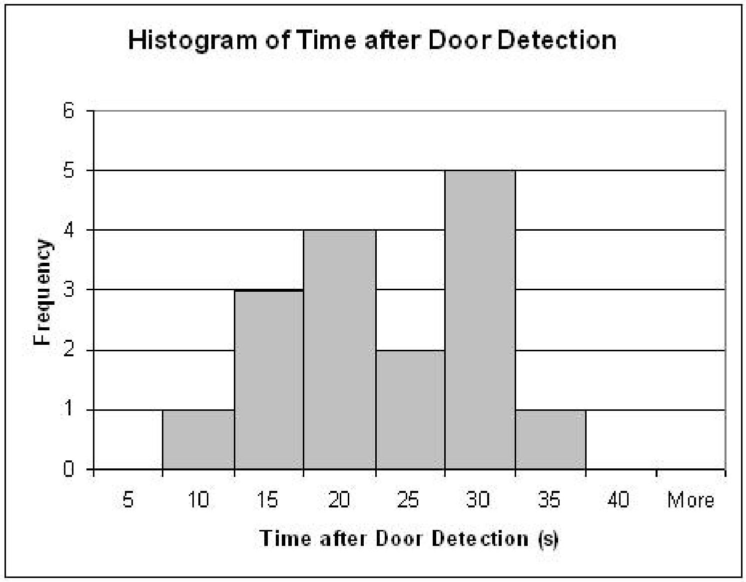


Figure D.2: Distribution of times after door detection: A relatively stable amount of time required to travel to and open the door was observed. As users became more familiar with the system, times after door detection decreased according to their ability to use the system.

D.2 GPS and INU Mid-Range Outdoor Navigation Data

In the following figures, the dashed blue lines indicate the intended path of the blindfolded test subjects. The goal was to stay within a 10-foot-wide band. The red lines show the users' actual paths. Measurements were only taken when users deviated from the 10-foot-wide allowable path, so as long as they were within the path, their location was recorded at the x axis.

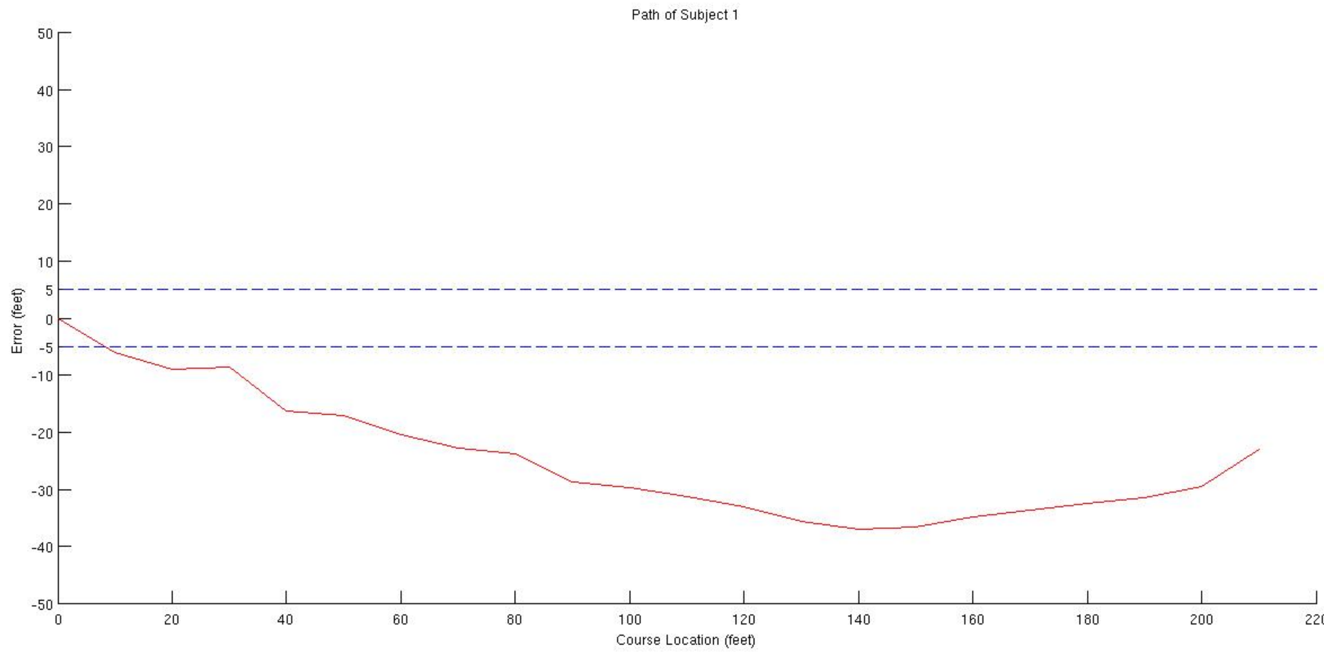


Figure D.3: Outdoor navigation path for subject 1

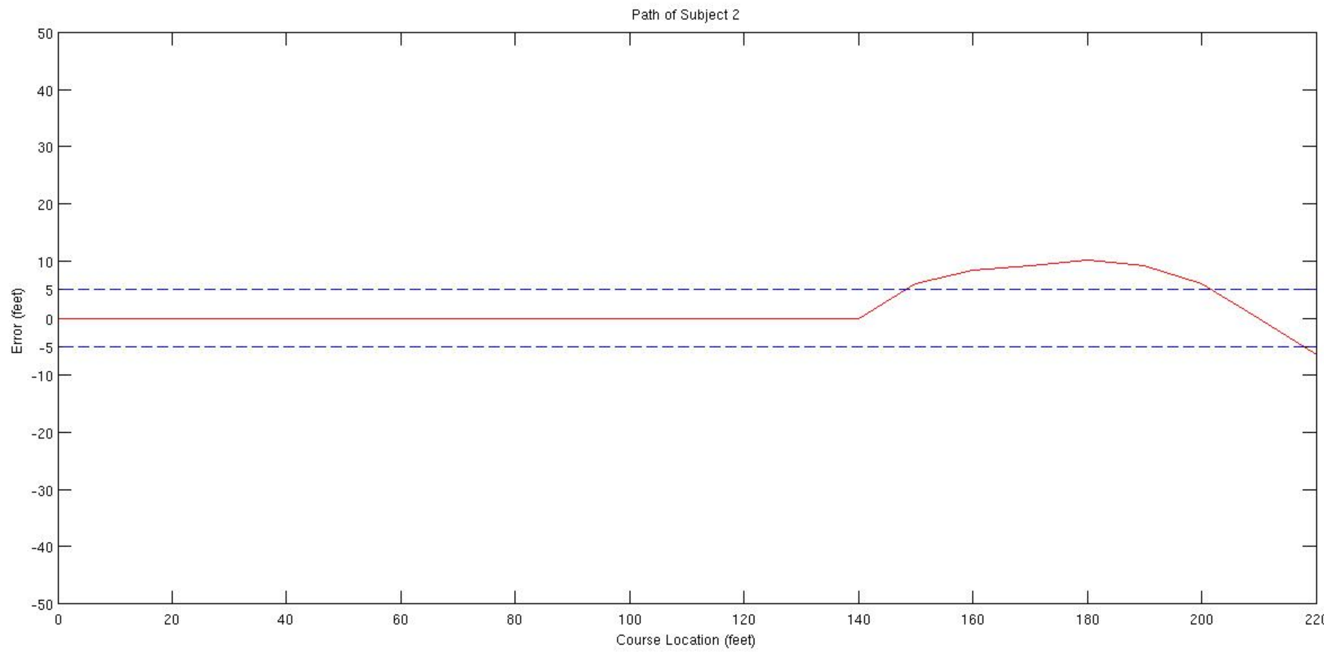


Figure D.4: Outdoor navigation path for subject 2

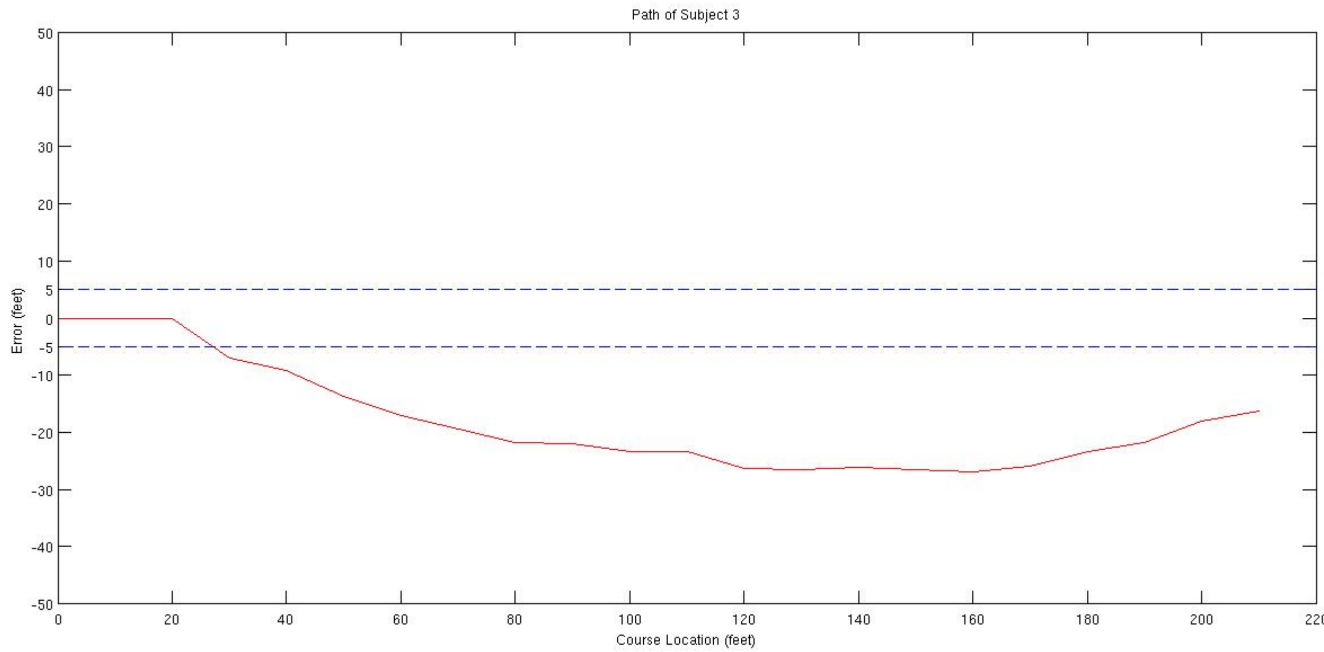


Figure D.5: Outdoor navigation path for subject 3

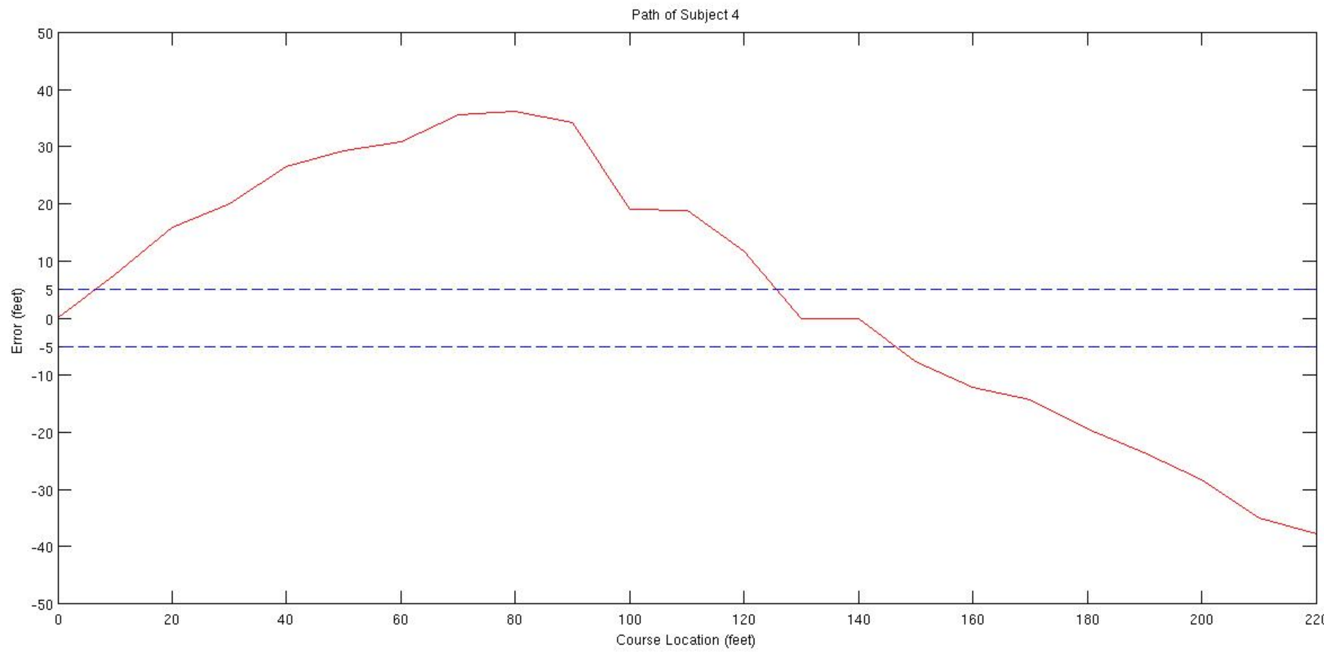


Figure D.6: Outdoor navigation path for subject 4

BIBLIOGRAPHY

[1] National Eye Institute. (2006) Progress in eye and vision research 1996 - 2006. [Online]. Available: http://www.nei.nih.gov/strategicplanning/NEI_ProgressDoc.pdf

[2] J. Borenstein and I. Ulrich, "The GuideCane – A Computerized Travel Aid for the Active Guidance of Blind Pedestrians," *Robotics and Automation, IEEE International Conference on*, pp. 1283–1288, 1997.

[3] L. Ran, S. Helal, and S. Moore, "Drishti: An Integrated Indoor/Outdoor Blind Navigation System and Service," *Pervasive Computing and Communications, Second IEEE Annual Conference on*, pp. 23–30, 2004.

[4] P. Ponchillia, N. MacKenzie, R. Long, P. Denton-Smith, T. Hicks, and P. Miley, "Finding a Target with an Accessible Global Positioning System," *Journal of Visual Impairment and Blindness*, vol. 101, no. 8, p. 479, 2007.

[5] L. Merabet, J. Rizzo, A. Amedi, D. Somers, and A. Pascual-Leone, "What blindness can tell us about seeing again: merging neuroplasticity and neuroprostheses," *Nature Reviews Neuroscience*, vol. 6, pp. 71–77, 2005.

[6] D. Ross and B. Blasch, "Development of a Wearable Computer Orientation System," *Personal and Ubiquitous Computing*, vol. 6, no. 1, pp. 49–63, 2002.

[7] J. Loomis, Y. Lippa, R. Klatzky, and R. Golledge, "Spatial Updating of Locations Specified by 3-D Sound and Spatial Language," *Learning, Memory*, vol. 28, no. 2, pp. 335–345, 2002.

[8] R. Datta, J. Li, and J. Wang, "Content-Based Image Retrieval - Approaches and Trends of the New Age," *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pp. 253–262, 2005.

[9] D. Lowe, "Object Recognition from Local Scale-Invariant Features," *Computer Vision, IEEE International Conference on,*, vol. 2, no. 2, p. 1150, 1999.

[10] S. Zickler and M. Veloso, "Detection and Localization of Multiple Objects," *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, pp. 20–25, 2006.

[11] K. Lyons and T. Starner, "Mobile capture for wearable computer usability testing," *Proceedings of IEEE International Symposium on Wearable Computing (ISWC 2001)*, 2001.

[12] N. Bradley and M. Dunlop, "An Experimental Investigation into Wayfinding Directions for Visually Impaired People," *Personal and Ubiquitous Computing*, vol. 9, no. 6, pp. 395–403, 2005.

[13] M. Tscheligi and R. Sefelin, "Mobile navigation support for pedestrians: can it work and does it pay off?" *interactions*, vol. 13, no. 4, pp. 31–33, 2006.

[14] M. Kolsch, R. Bane, T. Hollerer, and M. Turk, "Multimodal interaction with a wearable augmented reality system," *IEEE Computer Graphics and Applications*, vol. 26, no. 3, pp. 62–71, 2006.

[15] J. Marston and R. Church, "A relative access measure to identify barriers to efficient transit use by persons with visual impairments," *Disability and Rehabilitation*, vol. 27, no. 13, pp. 769–779, 2005.

[16] J. Marston, "Spatial knowledge acquisition: using technology, training, and techniques to enhance spatial learning for two special populations," *Cognitive Processing*, vol. 7, pp. 170–170, 2006.

[17] A. Krause, A. Smailagic, and D. Siewiorek, "Context-Aware Mobile Computing: Learning Context-Dependent Personal Preferences from a Wearable Sensor Array," *IEEE Transactions on Mobile Computing*, vol. 5, no. 2, pp. 113–127, 2006.

[18] R. Golledge, R. Klatzky, J. Loomis, J. Speigle, and J. Tietz, "A geographical information system for a GPS based personal guidance system," *International Journal of Geographical Information Science*, vol. 12, no. 7, pp. 727–749, 1998.

[19] S. Bains, "Mixed Feelings," *Wired*, vol. 15, no. 04, 2007, http://www.wired.com/wired/archive/15.04/esp.html?pg=1.

[20] F. Liarokapis, "Location-based Mixed Reality for Mobile Information Services," *Advanced Imaging Magazine*, April 2006.

[21] SiRF Technology Inc., "SiRFstarIII GSC3e/LP & GSC3f/LP," http://www.sirf.com/products/GSC3LPProductInsert.pdf.

[22] E. Dijkstra, "A note on two problems in connexion with graphs," *numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[23] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[24] J. Coughlan, R. Manduchi, and H. Shen, "Cell Phone-based Wayfinding for the Visually Impaired," *1st International Workshop on Mobile Vision, in conjunction with ECCV*, 2006.

[25] J. Coughlan and R. Manduchi, "Functional Assessment of a Camera Phone-Based Wayfinding System Operated by Blind Users," *Conference of IEEE Computer Society and the Biological and Artificial Intelligence Society (IEEE-BAIS), Research on Assistive Technologies Symposium (RAT'07)*, 2007.