

Towards Building Cyber-Human Systems
for Individuals with Visual Impairment

by

Devi Archana Paladugu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2016 by the
Graduate Supervisory Committee:

Baoxin Li, Chair
Terri Hedgpeth
Robert Atkinson
Erin Walker

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

A lot of strides have been made in enabling technologies to aid individuals with visual impairment live an independent life. The advent of smart devices and participatory web has especially facilitated the possibility of new interactions to aide everyday tasks. Current systems however tend to be complex and require multiple cumbersome devices which invariably come with steep learning curves. Building new cyber-human systems with simple integrated interfaces while keeping in mind the specific requirements of the target users would help alleviate their mundane yet significant daily needs. Navigation is one such significant need that forms an integral part of everyday life and is one of the areas where individuals with visual impairment face the most discomfort. There is little technology out there to help travelers with navigating new routes. A number of research prototypes have been proposed but none of them are available to the general population. This may be due to the need for special equipment that needs expertise before deployment, or trained professionals needing to calibrate devices or because of the fact that the systems are just not scalable. Another area that needs assistance is the field of education. Lot of the classroom material and textbook material is not readily available in alternate formats for use. Another such area that requires attention is information delivery in the age of web 2.0. Popular websites like Facebook, Amazon, etc are designed with sighted people as target audience. While the mobile editions with their pared down versions make it easier to navigate with screen readers, the truth remains that there is still a long way to go in making such websites truly accessible.

This dissertation introduces several innovative end-to-end prototype systems that benefit from intelligent server side processing capabilities and blind-friendly

client side interfaces, all targeted addressing key needs of individuals with visual impairment in the era of smart devices and social media.

DEDICATION

To Harsha, my parents and Dr.Li. Thank you for the support, encouragement and
patience.

ACKNOWLEDGMENTS

My efforts were partially supported by two National Science Foundation grants (#0845469 and #1135616). Any opinions, findings, and conclusions or recommendations expressed in this material are those of mine and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
CHAPTER	
1 INTRODUCTION	1
2 BUILDING AN ONLINE COMMUNITY FOR SPECIAL NEEDS.....	5
2.1 Problem Statement.....	5
2.2 Case Study	8
2.2.1 Study 1: Need for a Community.....	8
2.2.2. Study 2: Understanding Sitemap and Layout.....	9
2.3 Proposed Design	11
2.3.1 Services to Provide.....	12
2.3.2 Work and Study	13
2.3.3 Daily Living	14
2.3.4 Navigation	14
2.3.5 Resources	15
2.3.6 Forum	15
2.3.7 Design of the Site	16
2.3.8 Customization of Map API	17
2.3.9 Architecture of the site	18

CHAPTER	Page
2.3.10 Forum Design	19
2.4 Experiments and Analysis	20
2.4.1 Qualitative User Feedback	20
2.4.2 User Activities	22
2.4.3 Add-ons	22
2.4.4 Customized Directions	23
2.4.5 Dynamic Pages	23
2.5 Sustainability	24
2.5.1 Scope of the Problem	24
2.5.2 Proposed Solution.....	25
2.5.3 Evaluation of Proposed Model.....	30
2.6 Conclusion	32
3 NAVIGATION FOR OUTDOOR SHOPPING COMPLEXES	33
3.1 Problem Statement.....	33
3.2 Proposed Solution	36
3.3 Design Guidelines for Building Apps for OSC.....	38
3.4 Tiered Information Creation	39
3.4.4 Interface Design	42
3.5 Experiments and Analysis	45
3.5.1 User Feedback on Overall System Design.....	45
3.5.2 Testing the Proposed Functionalities	46
3.5.3 Evaluating Navigation Support	46

CHAPTER	Page
3.6 Sustainability	47
3.6.1 Front End Interface	48
3.6.2 Evaluation	56
3.6.3 User Feedback	59
3.7 Conclusion.....	60
4 A VIRTUAL ORIENTATION & MOBILITY INSTRUCTOR.....	61
4.1 Problem Statement.....	61
4.2 Proposed Approach	63
4.2.1 Interface Design	64
4.3 Experiments.....	68
4.3.1 Participants	68
4.3.2 Feedback from O&M Instructors	69
4.3.2 User Experiments	72
4.3.4 Results	73
4.4 Conclusion.....	76
5 TRANSLATING FLOOR PLANS INTO VERBAL DESCRIPTION	77
5.1 Problem Statement.....	77
5.2 Relevant Work	81
5.3 Proposed System.....	88
5.3.1 The Browser Interface	91
5.3.2 The Landmark Generation Module.....	95
5.3.3 The Web Crawler Module	96

CHAPTER	Page
5.3.4 The Visual Extraction Module	97
5.3.5 Content Identification Module	97
5.3.6 Sentence Ordering Module	98
5.3.7 Sentence Generation Module	98
5.4 The Visual Extraction Module	99
5.5 Content for Verbal Description.....	101
5.5.1 Deficit with Existing Models	102
5.5.2 Learning from our Target Audience	103
5.5.3 Case Study Findings	106
5.5.4 Paradigms for overall description	111
5.5.5 Lexical Choices for Describing the Concepts.....	114
5.5.6 Rules for Content Determination.....	115
5.6 Verbal Description Generation	116
5.6.1 An Overview	116
5.6.2 Overall Geometry Understanding.....	117
5.7 Evaluation.....	123
5.7.1 Results	130
5.7.2 Discussion.....	143
5.7.3 Use Cases.....	144
5.7.4 Limitations	145
5.8 Conclusion.....	152
6 CONCLUSION.....	154

CHAPTER	Page
REFERENCES	157

LIST OF TABLES

Table	Page
1 Statistics Collected from Amazon Usage.....	9
2 Statistics Collected from Facebook Usage.....	11
3 Statistics Collected from GoingEasy.org	22
4 Sample Keyword Clusters Using Our Method.....	30
5 Background Details of Our Participants	125

LIST OF FIGURES

Figure	Page
1 An Overview of the Architecture of the Server Hosting the Methods and User Interaction.....	2
2 The Login Page	12
3 Functionalities Provided.	13
4 The Resources Page	15
5 Illustration of Customization of the Homepage.....	17
6 Navigation Page for Low Vision User.....	18
7 Navigation Page for the Blind User.....	18
8 Navigation Results Presented to a Blind User.	18
9 A User Friendly Forum.	20
10 Illustration of Keywords and Topics.....	27
11 A Baseline HDP Model Without Taking User Distinction into Consideration.	29
12 A Three Layer HDP Introduced to Model User Related Topics.	30
13 The Tiered Information Representation Scheme Used to Support Application.	39
14 An Overview of iExplore's Homepage and a Sample Output for Each of the Proposed Functionality.....	44
15 Home Screen of Our App.....	49
16 The User Side of Navigate.....	50
17 Server-side Interface with Mandatory Checklist. Volunteers See this After Accpeting a Request from a VI User.....	52

Figure	Page
18 From Left to Right. Static Map View Centered at the Shop Location. Detected Shop Span. Red Circles Represent Detected Corners. Detection of Different Types of Roads from Google Maps.	56
20 An Overview of Our Application.	65
21 A Sample Where am I.	66
22 An Overview of the Steps Involved in Training.	66
23 An Overview of the O&M Training that We Could Implement as an App.	73
24 The Overall Architecture of Proposed System	89
25 Shown to the Left is the Website of a Public Building. The Red Circle Shows the Link that Redirects to the Image of the Floor Map. Shown on the Right : The Floor Map that the Link Leads to.	91
26 An Overview of the Workflow of the System.	93
27 Input Interface. (a) The Entry Point for the User. (b) Confirmation from the User. (c) The Final Output.	94
28 Workflow of the Data Processing Module.	99
29 A Few Sample Floor Plans Found on the Web Using Google Image Search.	110
30 "This Building is Rectangular in Shape which is Longer Across the Length or in the Horizontal Direction. The Main Entrance is Located at the Bottom Center of the Rectangle."	118
31 The Image Above Illustrates the Segments Used to Divide the Map.	119
32 We Identify Instances with Our Grid Pattern and Create the Corresponding Distance Information.	123

Figure	Page
33 The Materials Used in the Experiments.	127
34 A Selection of Results From a Subset of Users.	132
35 Some Sample Reconstructed Shapes Using Our System.	133
36 The User Recreated Maps from Cutouts and Verbal Description.	134

CHAPTER 1

INTRODUCTION

Until the last decade, solutions to aide individuals with visual impairment were in the form of expensive cumbersome equipment that needed to be acquired and carried around. With the advances in technology and wide spread popularity of hand held electronics, we look into developing solutions that are not farfetched and are accessible to all. Screen readers coupled with accessibility laws have made strides in making the Internet accessible. This made plethora of information that was previously transcribed tediously, available for everyday use. Increased connectivity and low cost devices makes this information travel friendly. While accessible technology has been playing catch up, the advent of web 2.0 changed the way we interact and access information. Ever changing content and layouts present the previously presented information in customizable formats. This presents advantages as well as disadvantages to individuals with visual impairment. My research studies some of these issues and attempts to propose methods to bridge the gap.

First issue we addressed is looking into the problems faced with web 2.0. A perfect example to illustrate this is the Facebook news feed. The feed looks different for each login. The backend recommendation algorithms are in play to customize and present content based on previous evident interest. While these techniques are optimized for presenting information visually, accessing this information using a screen reader makes it harder to navigate. We conducted a case study with 5

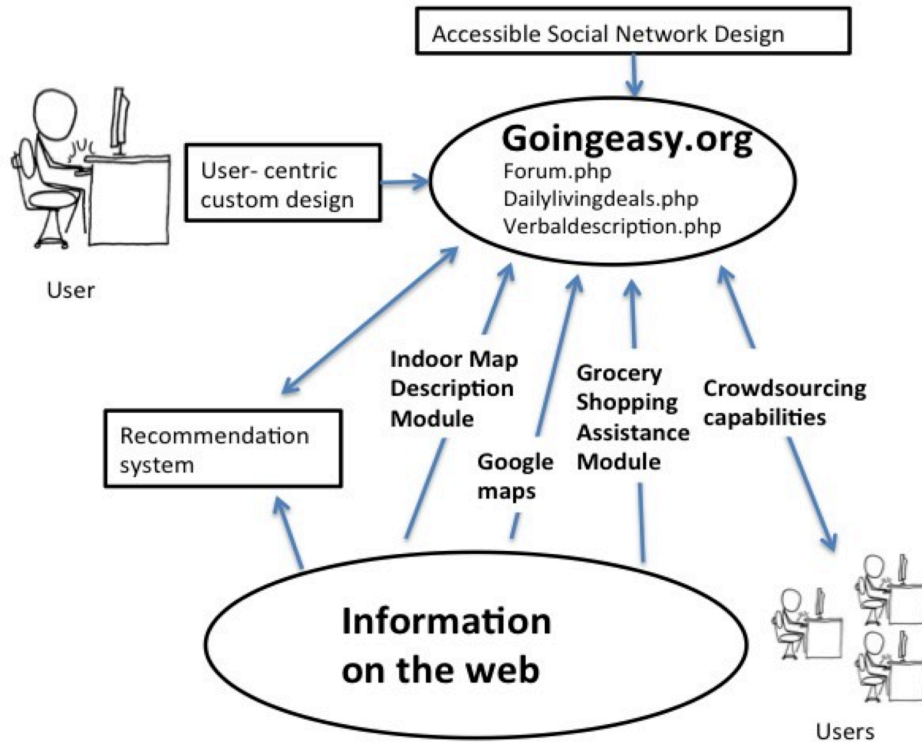


Figure 1 : An overview of the architecture of the server hosting the methods and user interaction.

participants to learn the problems faced on web 2.0. We looked into websites like Amazon, YouTube, Facebook and Orkut to understand more about navigating the web using a screen reader and also ways to design the layout of the websites make it screen-reader friendly. We developed the Goingeasy.org (<http://goingeasy.org/>), a social networking site focused on making information accessible to individuals with visual impairment. We then went on to develop modules that provide information like grocery store coupon information, accessible maps, discussion board, blogs, etc. We live in a picture dense Internet age and having these consolidated modules aim to reduce some of the effort spent searching for information.

Stemming from the lack of accessibility with respect to reference images – is a need for making maps accessible. Until recently, tactile maps or oral directions were the only way a map could be conveyed to an individual with visual impairment.

But smart interfaces have made it possible to reduce the dependence on asking for help. Apple's accessibility features like haptic feedback, speak screen, Siri, dictation, etc. on iOS made smart phones accessible for the first time making it possible to develop applications that aide navigation that can be accessed on location. We looked into the problems faced when navigating small range distances in outdoor locations like malls, down-the-street grocery stores, campuses, etc. From our surveys, we discovered that some critical information that is deemed essential like the presence of crossings, parking lots, obstacles and traffic conditions play a huge role in making a route safe. Such information is not readily obtainable. We propose ways to obtain additional information and integrate it into maps. We then discuss ways to present this information in a user centric manner. We evaluate our design principles by building an iPhone application that is built upon our case studies and findings.

The wide spread availability of outdoor maps in a standard format has made it possible to attempt making them accessible. Navigating close range spaces and indoors is still an area that hasn't seen practical widely implemented solutions. This is partly due the fact that blue prints for buildings are not something that is readily shared for security reasons. We attempted to use the power of semantic web mining to obtain the necessary blue prints and convert them to verbal description of the place intended by the user. In this process, we learn about the differences in information needed with respect to outdoor navigation, the terminology used and effective ways of communication when it comes to representing the information in question. We tested our prototype system with 19 individuals with visual

impairment from various walks of life. The feedback was indeed very promising and the users were able to reconstruct maps from the verbal descriptions.

With the above-proposed solutions (figure 1) , we tried to take steps towards building systems that help individual's aide independent lifestyle. These are more of engineering problems than research problems that are in dire need to a solution. We find ways to integrate image processing, pattern recognition, machine learning, field experience and interface design to create innovative solutions. We elaborate on each of these problems and proposed solutions in the subsequent chapters.

CHAPTER 2

BUILDING AN ONLINE COMMUNITY FOR SPECIAL NEEDS

2.1 Problem Statement

The social dimension in sharing of information has gained popularity over static sources of information on the web. Virtual communities, crowdsourcing websites and other information mining communities have become extremely popular method of information transfer. Eight of the top fifteen websites listed as most popular websites today listed by ebizmba.com are a form of a website that is sustained by user contribution. Existing web interfaces and designs are primarily designed, developed and tested with sighted users as target audience. The layouts, visually pleasing details, the graphics, the advertisements, the recommendation systems, etc are all designed for the sighted audience. The webmasters are often not fully aware of the challenges faced by the visually impaired users face in actively contributing and fully utilizing the underlying services. We believe exploring the differences and identifying potential changes to the design is a very important factor in making more accessible web 2.0 platforms in the future.

Significant work has been done in establishing guidelines for accessible webpages. In 1998, the United States Congress amended the Rehabilitation Act to require federal agencies to make their electronic information and information technology accessible to people with disabilities [1]. Similar legislation exists in Europe and a small number of countries. A decade later, studies show that disabled users spend almost twice the time online browsing webpages compared to their peers because of difficulty in navigation [2]. Statistics indicate that only 29.2 percent of blind/visually impaired people use the internet in comparison to sighted users

who stand at 59.7 percent. Accessibility is a subjective term and hence the guidelines are very hard to define and follow completely. Many online guidelines exist that give comprehensive suggested coding techniques to the webmasters. HearSay and Webvisum (www.webvisum.com) are examples of accessibility services that can be integrated into browsers. No work has been done on developing interfaces that address making optimal layout, organization and architecture of a blind-specific site. Companies like Google, Amazon, Facebook, etc have special accessible pages to reach out to a larger audience. But most of the accessible pages are concerned with proper HTML tags associated with the pages or scaling down of functionalities in a page to make the page less cluttered.

We explore the idea that “accessibility” can go beyond the practices of putting proper tags in the HTML code and de-cluttering visual content. Right tags and alternate text helped with the static web but the web 2.0 interfacing poses significant challenges. We consider and study issues related to the design of such a system, answering questions like: Can layout make a page easier to navigate? How can the architecture of the Website be made blind-friendly? Can an adaptive interface layout based on learning from user statistics enhance the usability of a page? What are the problems faced by independent websites that are separated from the mainstream social networks? What sort of recommendation systems work on these networks? These are the issues we address in the subsequent sections. We also present user studies, including those that were performed for establishing the design principles and those that were carried out for evaluating the pilot system.

The ultimate goal of our research is to develop a social networking site that enables visually-impaired people to form loosely-connected groups, actively

contribute their information and knowledge, and ask/answer unique questions that address special needs. This paper presents our initial design, implementation and evaluation of a pilot version of this site named GoingEasy® (www.GoingEasy.org). The current version of the system is intended for people in the same local area and for the users to contribute and share information that are directly relevant to their daily lives. For example, a user on GoingEasy® can ask for detailed walking directions for getting to a location, and he/she may get responses from other users. Such responses, when posted by a user with visual impairment who is familiar with that location, may be especially valuable since it can be very “blind-specific” and “blind-friendly”.

The latter part of this work explores the problems faced in sustaining this community. Since its content is generated centered around an interest/objective, user participation is very unlike a community like Facebook.com, where users typically log in multiple times a day and contribute on a very frequent basis. In such small communities, user contribution typically occurs in bursts, triggered by an interesting conversation or an event or some occurrence that triggers the members to log in at the same time frame. Members typically sign up because they are interested in the underlying cause or interest but fail to keep track of the information because they do not visit the website on a daily basis. But when they log in, they do participate and contribute. We use user participation problem in the scope of users who are committed to the network but fail to contribute because they can't keep track of the content.

We try to answer the following questions: What are the common problems faced by this community with respect to sustaining participation? Does it face

additional problems due to its primary audience? Do traditional solutions work? In this paper, we propose a recommendation system an add-on to the existing online community, to promote user contribution and sustain user interest.

2.2 Case Study

We conducted two phases of field studies when we started exploring the problem in hand. We had 5 users in our focus group studies: 3 totally blind and 2 low-vision users. These participants are of different academic background, age, number of hours of Internet use per week, proficiency with the screen readers, and types of blindness. The details of the studies are given below.

2.2.1 Study 1: Need for a Community

The first phase was intended to understand user concerns on information on areas where information deficit is felt the most and examine the need for a new interface. The three areas where the information deficit was felt the most were identified to be in the areas of navigation, daily living information (e.g., grocery deals), and information retrieval. All the participants confessed to having trouble in at least 2 of the 3 mentioned areas and they would resort to Internet search as a means to finding ways around it. They also brought up the concept of “blind-specific” daily living information, which is illustrated by the following example. The users were asked to name some local places that were not blind-friendly. One user immediately named a very popular shopping mall frequented that she deems extremely disorienting to blind costumers since the mall has multiple speakers playing music in the parking lot, causing the blind user to mistaken the parking lot for the stores in the mall. This is one of the reasons one of the existing social

networks is not a solution. The users specified that they wouldn't want to write such posts on Facebook because it's too invasive.

Table 1. Statistics collected from Amazon usage

Participant	Level of familiarity of the user	Time taken to find an item on Amazon	# of times assistance was asked
1	Used once or twice	3'10"	2
2	Never used	6'10" (fail)	1
3	Familiar	2'50"	1
4	Never used	4'20"	0
5	Every day user	2'11"	0

2.2.2. Study 2: Understanding Sitemap and Layout

The second phase of the study furthered our understanding of accessibility, usage patterns, and common problems. In this phase, the users were asked to perform 7 tasks (like finding a friend on Facebook, etc). Their usage pattern was recorded which helped us during the design stage. They were also asked 35 questions. Most of the questions were chosen from the set recommended in [3]. We also composed additional questions to collect the user ratings, timings on completing different basic functionalities on some popular websites – Amazon, Amazon Access, Facebook, Craigslist, YouTube and our site GoingEasy®. We collected statistics on usage of the sites by the participants, common problems faced, opinions on what they thought would be better. We found that that navigating the main-stream websites is still very challenging for the visually impaired. For example, the median

time needed for the users for joining a group on Facebook is 2'42" (compared to merely about 10" by an average sighted user). The reasons for the large amount of time taken to perform these tasks include the linearity associated with the screen reader, accessibility violations in the page design, visually pleasing design details of the pages that are hard to navigate, etc. We also tested different layouts, observed usage of functionalities of the sites, and analyzed the collected information as shown in Table 1. It can be seen from the table that navigating the main-stream websites is still very challenging for the visually impaired. For example, the median time needed for finding an item on Amazon for our users is 3'10" (compared to less than 30" by an average sighted user), and the median time needed for the users for joining a group on Facebook is 2'42" (compared to merely about 10" by an average sighted user) as shown in Table 1. The reason for the large amount of time taken to perform these tasks is the linearity associated with the screen reader, accessibility violations in the page design, visually pleasing design details of the pages that are hard to navigate, etc. Also, when a page is divided into panels with links grouped on each panel, it is hard to predict the order in which the screen reader accesses the content. The users who learnt some advanced shortcut keys in the screen reader software could get around faster. And this number was very small when compared to number of users we surveyed. The field studies provided a lot of important observations that were later reflected in our design of the GoingEasy® site. At high level, two critical design considerations are to make the layout blind-friendly (supporting easy and efficient navigation) and to include functionalities/information sources that keep people motivated to use and contribute to the network. These will be elaborated in the next section.

Table 2 : Statistics collected from Facebook usage

Participant	Level of familiarity of the user	Time taken to join a group on Facebook	Number of times assistance was asked
1	Familiar	2'00"	0
2	Familiar	1'55"	0
3	Very Familiar	2'56"	1
4	Never used	6'33"	2
5	Used once or twice	2'42"	0

2.3 Proposed Design

We reached our initial design based on the field studies described above and updated the solution through multiple iterations. The first version of the proposed site, GoingEasy.org, is a social networking website that was built with visually-impaired and blind users as the targeted audience. A major force driving social networking websites is having something unique to offer that is necessary and unique. It is worth noting that GoingEasy is designed with blind users in mind (WAI compliant), not built to be visually pleasing alone. GoingEasy® was built using PHP, JavaScript, C++ and MySQL. It requires sign-up so as to support custom user modeling of the pages. When a user registers on the website, he indicates his preferences, disability type, location, etc. The current pilot implementation of the site reflects our observations and conclusions from the field studies presented above. In this section, we elaborate the salient design features. As an introduction, we show the homepage of the GoingEasy® site in Figure 2. A registered user can log into the system, while a new user can register an account to begin with.



Figure 2 : The login page

2.3.1 Services to Provide

Conventional wisdom is that crowdsourcing networks are hard to sustain without monetary incentives involved [4]. We approached this problem with coming up with services that are deemed essential and were requested from us by the end users in our field studies. As discussed before, we observed that navigation, daily living information (groceries, entertainment) and specific information retrieval are the major problems faced by the visually impaired users in their everyday life. This, in conjunction with the requirement of being simple to navigate, helped us in defining the key functionalities (and interactive buttons) to provide in the first page of our site after a user logs into the system. Our design of this page is given in Figure 3. We would like to note that, while the page illustrates the key functionalities or information sources the site intends to provide, in the current version, some of the functionalities have not be fully loaded yet. However, keeping all the buttons enabled us to evaluate the fully-loaded functionalities with an

intended interface. We elaborate below what the functionality buttons are intended to provide.

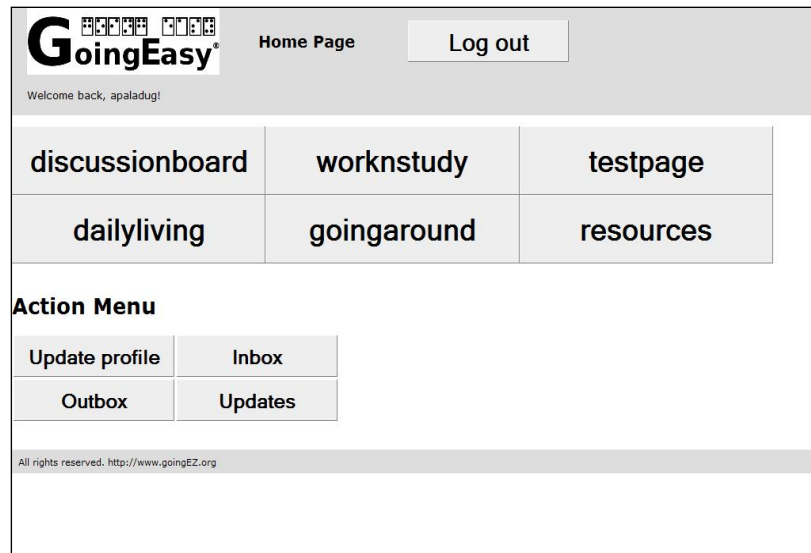


Figure 3 : Functionalities provided.

2.3.2 Work and Study

This page caters to information needs of the blind students and working professionals. We found that even in a very disability friendly campus, students and/or their guardians may still need to come to the campus' disability recourse center for information that is not available online. This is an example illustrated what the WorknStudy page may contain. We emphasize that this is different from a generic site such as the American Foundation for the Blind site (afb.org) because our site caters to a local community, and it is the local people who contribute to the information. Other topics to be included in this page include local jobs, conferences, places that convert books to alternate formats, blind activity clubs, etc.

2.3.3 Daily Living

We identified an area where the accessibility issues come in the way of making an informed decision. Most of the entertainment and coupon websites are rich in graphical content and not accessible. A sighted user has access to printed coupons, discount flyers, mail in coupons, emails rich with graphical description and prices of items, etc. Thus we have a dedicated Daily Living page that caters to providing information such as ads from local grocery stores and places of entertainment.

2.3.4 Navigation

This is an area where we identified the blind people to have most difficulty. Our navigation page is customized to the user depending on whether the user is sighted, low-vision, or blind. We offer customized settings for Google Maps and a provision to request for customized tactile maps for any destination in a reasonable radius. We customized Google API to read out text information and set the mode of transportation to walking under the assumption that the users are covering short distances on foot.

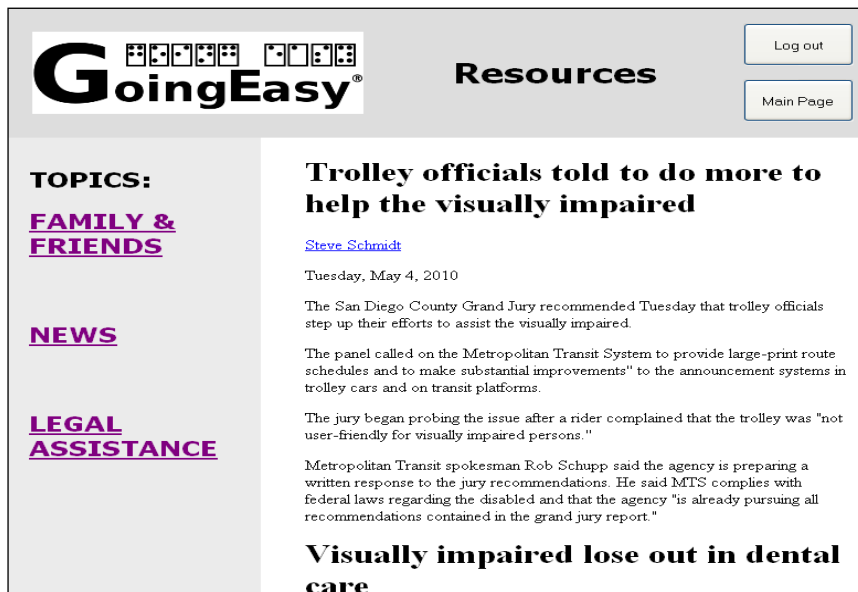


Figure 4 : The Resources Page

2.3.5 Resources

During our field study, participants confessed that their family did not know about the rights their blind kids had, the facilities that should be demanded from educational institutions by law, and some local resources that would have made a difference in their lives. We have a dedicated Resources page that collects and provides such information for the guardians and friends of visually impaired people. We also provide local and global news related to Visual Impairment/blind individuals to increase the awareness levels among the users. Figure 4 illustrates what a user may see from the current Resources page.

2.3.6 Forum

A forum was deemed essential for exchange of information, record experiences and as a medium for people to help each other out. It also provides a platform for crowdsourcing and active interaction. We have created a very basic clutter free forum with every confusing add-ons stripped off. A sighted person can

look at a forum and at a glance; infer the latest thread, most active thread and most posted thread. A visually impaired person would have to go through the thread topics sequentially to determine these parameters. We have provided options like sort by popularity, activity and default by timeline.

2.3.7 Design of the Site

The language used for our current development is HTML, PHP and JavaScript. The database was implemented using MySQL. After the initial rounds of field studies, the salient design features that have been included in the solution are presented below.

2.3.7.1 Dynamic Homepage

We explore ways in which the dynamism associated with the Web 2.0 technologies can actually be used to the advantage of the blind users. A sighted user glances at a page and can search for the link he/she is looking for in no time. But for a screen-reader user, she/he would have to go through all the links sequentially to get to the link she/he needs which is time consuming. (They could use the shortcut keys once they are familiar with the page and know what exactly they are looking for. But not many users in our surveys knew all the useful shortcut keys.) We propose to reduce this time by making customized dynamic pages that learn and adapt to the user behavior. Every time a user clicks on a link, uses a page or functionality, the user statistics are recorded. For example, if the user uses the forum a lot, and hardly uses the other functionalities in the homepage, that should be the first thing he should hear on the screen reader in his homepage. We currently sort out functionalities in descending order of usage.

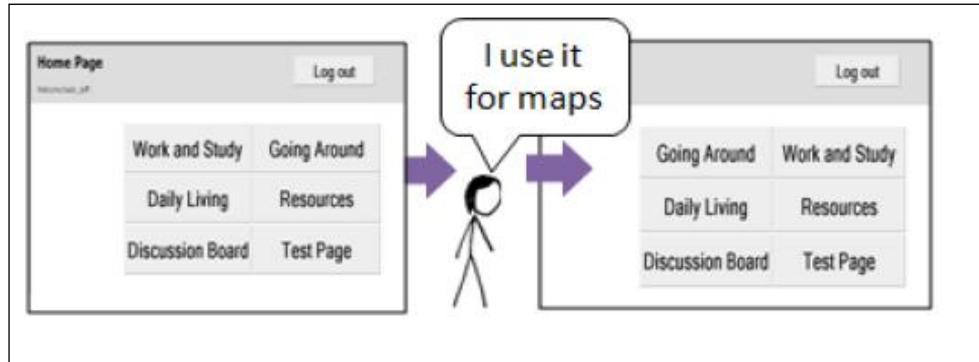


Figure 5 : Illustration of customization of the homepage

In our field study, we explicitly asked the following question: can a dynamic webpage help? All the younger users (college students) acclaimed the idea while their counterpart had reservation (and asked for an option to keep everything static). This has been reflected in our design, and is behind our emphasis on keeping a proper balance between dynamic layout for saving navigation time and static layout for supporting a sense of familiarity.

2.3.8 Customization of Map API

A salient feature we provide in the Navigation page is an interface customized to each user's needs. When a user creates his or her profile, if he chooses to indicate his visual disability type, the navigation page tailors itself to provide navigation instructions as her/his needs. For example, a blind user's page contains only text information about the directions. A low-vision user will see an enlarged version of the map and directions. A sighted user will see a map and instructions as would normally be returned by Google Maps. These are illustrated in Figures 5 to 7.



Figure 6 : Navigation page for Low Vision user

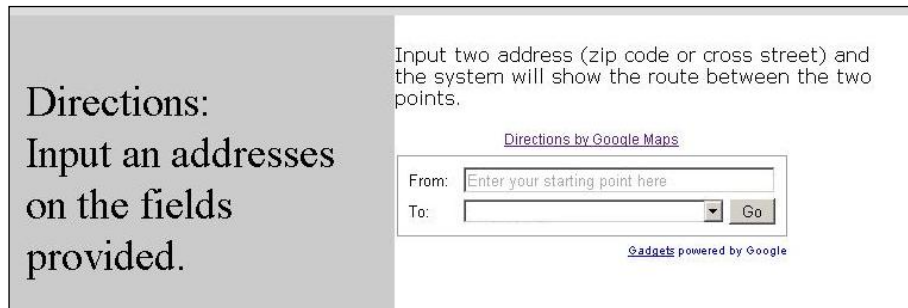


Figure 7 : Navigation page for the blind user.

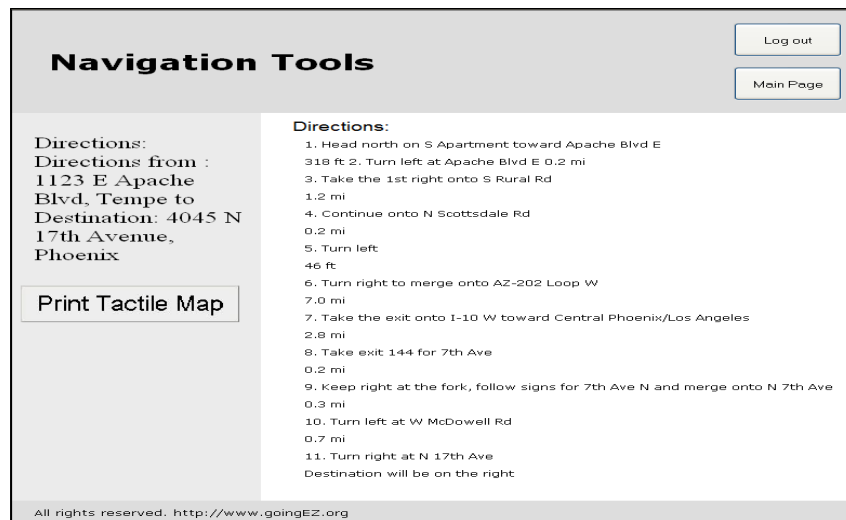


Figure 8 : Navigation results presented to a blind user.

2.3.9 Architecture of the site

Traditional social networking sites that we see have a homepage that contains a link to every functionality that the website offers. We observed that this

kind of architecture is what takes up a lot of time when trying to find something on a page by a screen-reading software. When the architecture of the site is changed to a more distributed tree like structure where the functionalities are hierarchically grouped, the user may be able to navigate to the desired button/function quicker. We tested two kinds of architectures for GoingEasy®. The first architecture contained all the functionalities in the site listed on the homepage like most existing sites like Amazon, Facebook, Orkut, MySpace, etc. The second one had a layered structure where similar functionalities are grouped and presented in each layer (with much few buttons) while the buttons can be expanded in the next layer. Every user preferred the layered structure better and disliked websites that were not tiered because the screen readers read out a lot more information every time they visited the page. This has been reflected in our design of the homepage and the pages for the functionalities (Figure 3).

2.3.10 Forum Design

The forum is a critical part of the design that aids in collecting information from the users. A screen reader scanning the forum sequentially becomes a disadvantage and time consuming as the size of the forum increases. A blind friendly forum should include a way to sort the forum as per user preference. GoingEasy® has a default view of the forum as well as options to sort the forum based on activity, popularity and time. And the forum is very clutter-free and easy to use, requiring minimal learning. The users disliked the idea categorizing topics in the forum and hence this detail was removed. Figure 9 illustrates a view of the current forum design.

Sort by Latest Post	Sort by Popularity	Sort by Timeline	Main Page
List of topics in the forum:			
TOPIC TITLE		Number of Posts	
Ipad and iphone Created on May 3 2010 at 11:55:24 PM by apaladug		3	
Complaints to the web administrator Created on May 3 2010 at 05:19:26 PM by apaladug		1	
Ask for Information Created on May 3 2010 at 10:46:34 AM by apaladug		7	

Figure 9 : A user friendly forum.

2.4 Experiments and Analysis

The main features to put forth in this paper are a crowdsourcing community as a solution to meet information deficit, add-ons to motivate people to use and contribute to this network, usage of dynamic pages to eliminate the linearity of a screen reader, and user input based customized maps. The current version of the system has implemented all these features and in this section we report the evaluation of the system based on three users, two of which are blind and the third is low-vision.

2.4.1 Qualitative User Feedback

The users were asked to take a survey about their experiences with using GoingEasy®. This survey was designed using criteria from (<http://www.pages.drexel.edu/~dea22/questions.html>). All the users mentioned the effortless in signing up to use the website. They praised the effort for enabling them to connect to other visually impaired people on common grounds and not having to worry about accessibility issues. All of them indicated that they faced no webpage accessibility issues in accessing GoingEasy®. The forum initiated a

discussion on a wide variety of topics like useful websites, information on a shopping malls, entertainment related information about movie theaters, support and encouragement for a new user, appreciation of the local heroes, discussion of technical products, recession in the US, etc. We observed that GoingEasy® promoted a sense of community among the users. The users indicated a strong comfort zone when interacting with the fellow people in the network because they have the same issues either knew or heard about some of the people they were interacting with and were talking about issues in the same locality as each other. Users also expressed using the forum for altruistic reasons and felt good about helping each other. The users expressed a liking for the platform and wanted us to scale it up to make it better. We quote a participant's comments below from the user evaluation:

“I would like to see this broadened to include blind people both in the city of Phoenix and the state of Arizona. I think the more people that sign up for the service, the livelier the discussions could become. Most discussion for the blind are email messages sent back and forth between individuals or through listservs. Having a website where messages could be sent and answered almost would, I think, be a godsend.”

According to [5], the quality of a network can be characterized by its ability to connect with people in new ways: The effortless to sign up, its ability to shift power from institutions to the people, generation of enough content in the community to sustain itself and having an open platform that invites partnership. Based on our current study, we believe that the current version of GoingEasy®, while with only limited functionalities in every aspects, indeed promotes the first four of the five mentioned criteria to make a successful social network.

2.4.2 User Activities

The forum and the statistics from the backend database are the indicator of usage of website. The statistics indicate that the registered users visited the site on an average of 5.5 times per day. The users reported spending anywhere around 10 to 30 minutes per day browsing the content or the forums on the website. Table 3 summarizes such data. The topics discussed covered a wide range of aspects and instigated very helpful posts. When we told them this was a testing prototype, all the participants requested us to continue with this service and asked us to advertise this service on major disability conferences.

Table 3 Statistics collected from GoingEasy.org

Participant	Level of familiarity of the user	Time taken to find a certain article	Number of times assistance was asked
1	1	<30 sec	0
2	1	<30 sec	0
3	1	<30 sec	0

2.4.3 Add-ons

The Navigation (GoingAround), Resources, Daily living, and Work and study pages were introduced on request of the users. And the users helped decide the content to put in these pages, and worked to make them useful. This did promote a sense of ownership on the content and had a higher scale of responsibility and awareness to contribute. Some of the content on the forums made it into the static pages with the users being the contributors. While currently such inputs from the users were directly collected from the users in person since we had only a small number of participants, the site has the potential of gathering such inputs

automatically through the forum from a larger user group in the future. Thus we expect that the above benefits will still be felt by future users.

2.4.4 Customized Directions

Every user in our experiments got to see a navigation page that was relevant to him/her. The users gave us feedback saying they were happy to hear textual directions that were walking directions from a source to a destination. And the users could obtain tactile maps of places which they said was indeed novel and very convenient. Most of the users requested maps of where they lived to learn about their surroundings better.

2.4.5 Dynamic Pages

We introduced an adaptive homepage tailored to user statistics. In the phase one field study, participants replied affirmatively when we explained the idea. But some of the users had apprehensions because every time a page changes, it would involve some getting-used-to period. From the feedback we got from the users after they started using GoingEasy®, most of the users did not even sense the changes in the homepage. When they were later told about this feature, they were surprised that they did not notice the changes. But they did mention how “easy” it was to find what they wanted. They did not have a problem with the website maintaining statistics on their usage histories on the site. It helped our understanding of the navigation paths on the site and user behavior.

2.5 Sustainability

GoingEasy® can be classified as a small interest based community (SIBC). Its characteristic's are as follows: Its is concentrated around a locality(constraint imposed by the developers). Since its confined to visually impaired users, it has a very few users knit by a common interest and problems. This network is very different from Facebook.com, which has all of user's friends, and the interaction is very casual. Also, users don't log into the network every day or multiple times a day as is the norm with Facebook. GoingEasy® has people participating and contributing for altruistic reasons.

GoingEasy® faced gradual waning of user contribution. Upon being asked about the reduced frequency of posts, the users confessed to not wanting to log in everyday to check for posts. Also, they didn't like the fact that on days they did log in, there may or may not be a post that they cant relate to or interests them. We also observed lots of posts not having any responses. When asked about it, the users admitted to having not read the post. The proficiency in using the screen reader made a difference in overlooking some posts. We asked them if having a recommendation system like Facebook has, which sends a notice anytime any user writes a post. Most of the users vetoed the idea claiming that email-all emails are very bothersome and spam like.

2.5.1 Scope of the Problem

Many small virtual communities stop growing or fail because of the high turn over and not enough user participation. We looked at Facebook pages and groups to see how they encourage user participation. Facebook Users join many groups that

they seldom participate in. It can also be seen that many groups exist for a given interest. For example, a topic 'Running' has more than 100 public Facebook pages, out of which it can be seen that only 20 are active and upload content and promote discussion on a regular basis. We observed a Facebook page called Seeitourway.org and a Facebook group called Ability Counts, since our users mentioned they use these two communities to promote similar discussions as on GoingEasy®. It can be inferred from these communities also face similar problems. They also have users who join the network but do not contribute in spite of using Facebook on a regular basis. Facebook encourages user participation by giving out a notification to every post created to everyone in the group. It can be seen that this is not very effective and many times be overwhelming. In a visually impaired online community, overlooked posts are another major problem. It is not uncommon to miss certain posts when using the screen reader to skim through the content. More often than not, when using a screen reader that sequentially reads out the content, if the user breaks at a point, it is difficult to return to the exact same point and not overlook some content. These observations highlight a need for a more customized recommendation system in place, even for Facebook.

2.5.2 Proposed Solution

We propose a customized recommendation system that notifies a user when topics that he has expertise in or is interested in are being talked about on the discussion board. It would overwhelm a user who doesn't log into a SIBC on a regular basis to receive notification via email every time anyone posts some content. A good recommendation system customizes to the content posted and we believe can be used

to enhance user commitment and participation. Users can visit the site when their contribution is needed or interests are matched, based on the alerts given by the recommendation system.

2.5.2.1 Integration of Social Psychology Theories

Frazan et. al. [6] put forth research proving that motivators of individual effort in collective effort situations provide great benefit in increasing the participation in a community. Social Psychologies refer to this phenomenon as social loafing, where users exert less effort on a collective task than they do on a comparable individual task. They put forth a hypothesis that users contribute more when the uniqueness and the benefit of a users contribution are enunciated and highlighted. They built a recommendation system that sent an email to the users acclaiming their contribution whenever they rated a movie that had very few rating on their movie rating website. They also sent an email to this user when ratings are needed for unique movies. They also highlighted the benefit of each users contribution to the user. We built the foundation of our recommendation system upon this hypothesis.

2.5.2.2 User Interest Identification

The first step towards building this recommendation system is user expertise/interest identification. Topics of interest of the user are not explicitly stated but needs to be inferred from the past user history, which helps build a dynamic model. Also, the areas of interest are not obvious from a few posts on Facebook or GoingEasy®. A good solution to this problem is latent topic modeling. Latent topic modeling has become very popular way of inferring latent clusters in

the last decade because of the flexibility of defining structures and the ease of extendibility to the user level. Latent Dirichlet Allocation [7] was first proposed by Blei et al. in 2003 for text modeling, and has been a popular choice ever since. This model can be extended to infer a topic-keyword model and user-keyword model given number of latent topics. The drawback of this approach is the necessity to provide the number of latent topics as the input to the algorithm. With the growing content and scope of keywords, this was a major disadvantage. This problem can be overcome by using Hierarchical Dirichlet Processes [8], a very powerful non-parametric approach used for latent topic modeling. We propose to model the content on GoingEasy using probabilistic latent topic modeling as explained in the subsequent sections.

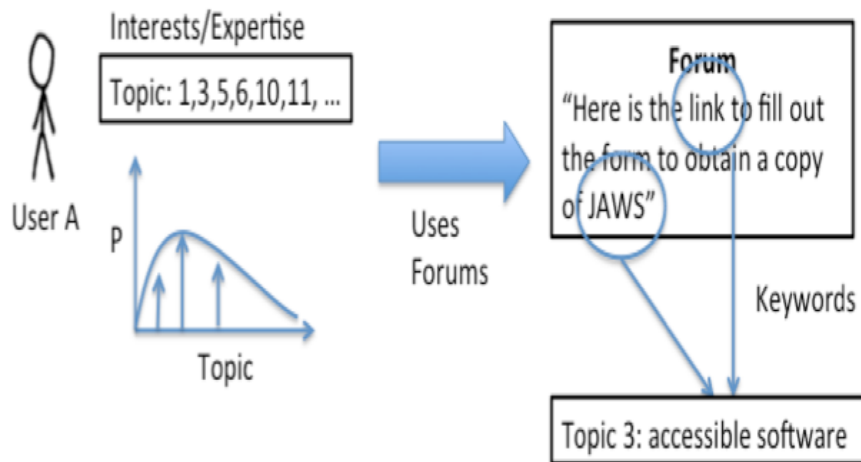


Figure 10 : Illustration of Keywords and Topics

2.5.2.3 Each Post as a Bag of Keywords

When a user posts a message on a forum, the post can be considered as a bag of keywords. Keywords can be extracted from a sentence using grammar-based rules, frequency based rules or occurrence based rules. Once the keywords are extracted, a post can be modeled as a mixture of latent topics and the topics can be

modeled as a mixture of keywords. Each post j can be assumed to be drawn from a model on certain topics.

$$G \sim \sum \beta_m \delta_m, \beta \sim \text{Stick}(\alpha)$$

where β defines the parameters of a stick breaking process and δ is an atom at each of the m latent topics inferred by the model. This model has the shortcoming of having no constraint to ensure that the similar keywords occurring in two different posts have same topic affiliation. This constraint can be enforced using Hierarchical Dirichlet Processes and using a base distribution that uses re-occurring atoms across the entire content corpus. We define a set of random probability measures $(G_j)_{j=1}^J$ where j denotes content related to each post. G_j has a concentration parameter α_j and a global random probability measure G_0 distributed as a Dirichlet Process with concentration parameter Υ and base probability H .

$$\begin{aligned} G_0 &\sim DP(\Upsilon, H) \\ G_j &\sim DP(\alpha_j, G_0) \end{aligned}$$

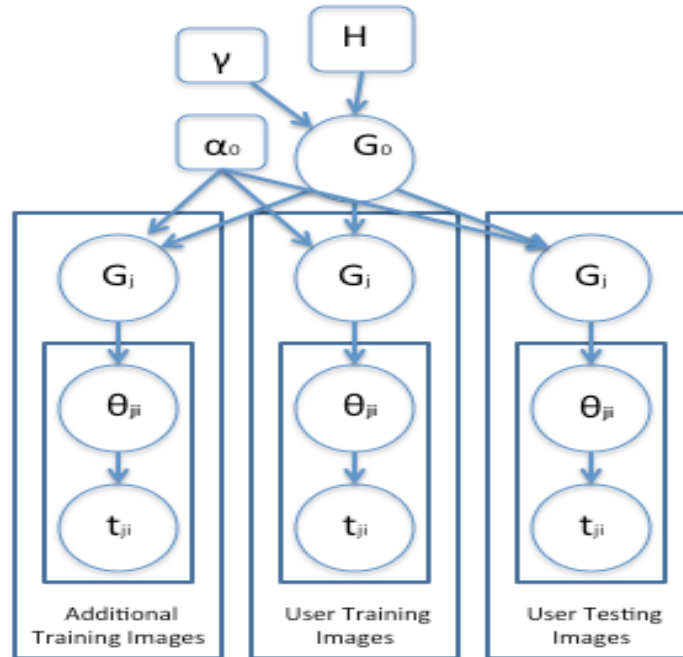


Figure 11 : A Baseline HDP model without taking user distinction into consideration.

A model described above fails to take user distinction into consideration when inferring the latent topics. This model can be extended to add another layer that models users as multiple groups. This layer captures the user-topic distribution from the corpus. We use Gibbs Sampling to evaluate a posterior distribution on θ and α instead of directly estimating the model parameters.

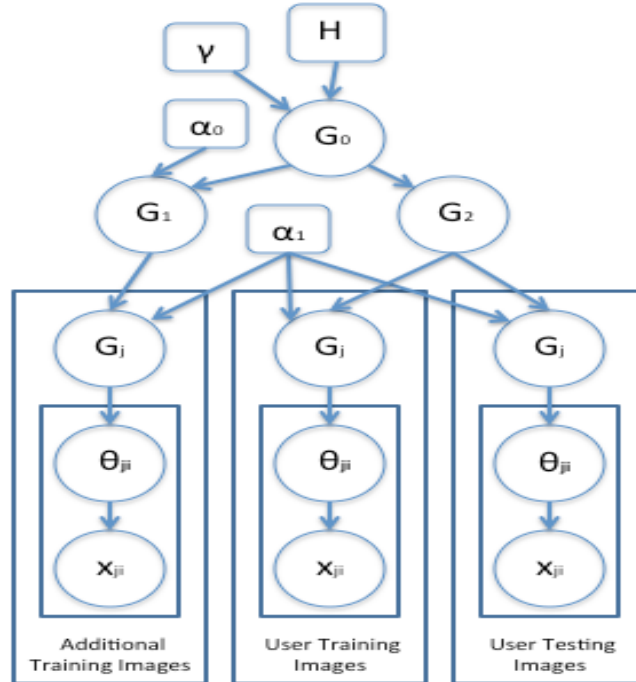


Figure 12 : A three layer HDP introduced to model user related topics.

2.5.3 Evaluation of Proposed Model

Evaluating a recommendation system is a hard task on a real time social network because it is hard to control other factors that are involved in a users response to a recommendation. We ran experiments with and without the recommendation system and we present our observations below.

Table 4 : Sample Keyword clusters using our method

	Topic 1	Topic 2	Topic 3
Keywords	Accessible Site Test Website Hotkeys	Jobs Proofread Braille Website access	Lightrail Stop Driver Contact Blind

We use the number of un-attended posts (i.e, a post created that did not catch any users attention and resulted in no answers), response rate derived from replies

to each post and number of recommendation notifications that lead to a reply post as the metrics to evaluate this proposed system. We considered two time frames for comparison, a peak one-week window before the introduction of the recommendation system (Phase 1) and a one-week window after the introduction of the recommendation system (Phase 2). GoingEasy® has a personal messaging system inbuilt. We used that to notify users to potential interest match. Users were asked to check their inboxes before they visited the discussion board.

When a user posts a new post on GoingEasy®, the model extracts keywords from the sentence and predicts topic(s) membership information. It then uses the stored topic-user relationship to predict the users who are most likely to respond to this post and sends them a notification. Since the predicted topics are in the form of a probability score, we associate each post to the top 2 topics. We then look at the members who have either of these two topics as their top interests in their user topic probability scores. We used the number 2 as the cut off threshold based on the number of topic and the users. This number can be varied as these numbers change. We chose email as the way to notify the users because of ease of implementation. Any other form of inbuilt notification system would be costly to implement and change at this stage.

As seen on the forum on GoingEasy, Phase one had 20% unattended posts and an average response rate of 3 replies per post. During the Phase 2 part of experiments, we observe no unattended posts and the average response rate increased to 6 responses per post. The conversion rate from the recommendation system is 76 percent, i.e when a recommendation was made, the users visited that link and contributed 76 percent of the time. We also used this model to re-kindle

some of the older posts on the forum based on potential interest match. Users visited not so new posts and continued discussion that was over a year old. This is a testament to what a simple recommendation system to do to help sustain a very small community. Users were asked to give us qualitative feedback at the end of Phase 2. They mentioned that they liked the fact that they don't get a notification for everything written in the form and liked the idea of tailored recommendation.

2.6 Conclusion

We presented a novel social networking site, GoingEasy®, which aims at providing crowdsourcing capabilities to help a local community of computer users with visual impairment for improved information sharing. We presented our design principles and implementation details. We also reported the findings from various phases of field studies in the process of developing the system. The evaluation results suggest that the proposed approach is very promising in providing a unique solution for meeting the information needs faced by the users with visual impairment. In particular, we found that the dynamic internet interfaces supported in our system was able to make information access easier without creating a burden of learning. The current version of EasyGoing® is currently a functioning version of the features presented.

CHAPTER 3

NAVIGATION FOR OUTDOOR SHOPPING COMPLEXES

3.1 Problem Statement

Navigating an unfamiliar outdoor space is a challenging task for visually impaired (VI) people. The use of technological devices in conjunction with specialized training and accessible architecture of buildings and roads has made this task possible. However, there are still many factors that contribute to making this task stressful and sometimes dangerous for VI users. Street layouts are The situation is worsened especially for outdoor building complexes such as outdoor shopping complexes (OSC), whose layout is often cluttered with haphazard placement of stores, parking spaces, walkways and driving roads, etc. Existing GPS devices seldom provide useful formation for the navigation of a VI user in such a setting.

The first problem towards building a system for OSC navigation is the lack of awareness among developers regarding the inadequacy of the current systems in place and lack of studies stating ways to overcome the existing information deficit. This paper reports our study on the challenges faced by VI users when navigating OSC and lists the necessary information required to bridge this gap. Secondly, the required information is currently not available from one source. We propose a technological solution for data fusion of information obtained from different sources. Thirdly, once the necessary information is in place, our application catering to VI users has presented a voice over for visual data. We evaluate these techniques and propose a system design that takes after the design paradigms obtained from our

user studies. This paper delves into these three aspects that are required for the development of this system.

The focus on OSC was motivated both by their direct relevance to the life of VI users and by the fact that conventional map services do not provide navigation for an OSC. Typical OSC are fairly complex in layout and thus serve well as representatives of general outdoor building complexes. Besides building a functional iPhone app that is freely available to our participants, the work contributes to developing general principles and guidelines in designing interface schemes and information representation paradigms for supporting mobile-device-based navigation assistance for people with visual impairment.

Navigation assistance for visually impaired users includes accessible infrastructure, specialized orientation and mobility training [9], and technological aids [10]. Technological aids, which are the focus of this work, may be devices that give live help on site or tools that help the user to prepare for the journey ahead of time. A review of dedicated devices for on-site navigation help using GPS, sonar, laser, RFID, etc. can be found in [11]. Recent years have seen new systems that are built upon general-purpose touchscreen mobile devices, moving away from the requirement of dedicated hardware. The most dominant players of this type are these three iPhone apps: Ariadne GPS (www.ariadnegps.eu), MotionX GPS (gps.motionx.com) and Lookaround GPS (www.senderogroup.com). Google's Intersection Explorer allows touch-based walk able path exploration. Recent years have seen add-on techniques to make the underlying maps more helpful by using audio, haptic [12] and spatial tactile feedback [13]. There are relatively fewer methods/systems for supporting exploration and spatial learning of locations (to help

a user's planning of a trip ahead of time). Examples include tactile map systems and verbal description generation methods [14].

Despite the existence of the aforementioned work, the reality remains that none of them can support navigation in an OSC setting largely due to the lack of adequate mapping information [15], even with the newer type of systems like Ariadne GPS. A solution to this deficit is to obtain information from various sources and fuse them before integrating them into the application. Some prior work on map registration and geo-spatial data conflation addresses the problem of combining data from different sources to obtain the information for the applications that require additional geo spatial information. A method was proposed in [16] to align successive images taken aerially with an overall map of the region using feature based registration and mosaicking techniques. Linear features like active contours to register images were proposed in [17]. A technique that uses feature matching using RMS estimation on affine transform was proposed in [18]. Street data were used as control points and then cross correlation was employed to obtain matching in [19]. None of these methods are directly applicable for the following reasons. Firstly, these methods have prior knowledge of the structure of their maps and available information in their maps. Our data fusion uses information obtained from standard maps as well as shopping directories crawled from the web which can be viewed as pseudo-maps. Secondly, detecting control points now becomes an altogether different task due to the dissimilarity between two images on a lower-level. The data fusion for our application needs to be flexible enough to work with a reasonable accuracy for images obtained online with large variation in their structure. Further, current design of some of those systems is essentially based on the concept of making the

underlying information (targeted at sighted users) more accessible to VI users, without taking into the real needs of the VI user in navigating a place.

3.2 Proposed Solution

The importance of having a system in place can be emphasized by the fact that Google and Bing maps are working on integrating floor plans of malls into their map service. Google Maps has a user interface [20] published for the user to identify three control points and scale the map on the top of their street view map for user input. Bing maps has also handled similar problem (<http://maps.google.com/help/maps/floorplans/>) without the need for human intervention on popular shopping malls. Still these map services are unable to provide navigation inside OSC which is our ultimate aim. Therefore there is a need for a framework with data fusion to truly build an application for a VI user.

We started an exploratory study with a group of visually impaired smart phone users in order to understand the challenges and identify deficits of existing solutions. Feedback from our participants and online surveys of visually-impaired communities helped us to conclude that that the iPhone appears to be the most used and preferred smart phone among users with visual impairment. Most of the users are familiar with the voice-over feature on the phone. Three of the five users we interviewed reported that they collect information from the store before planning their visit. The information asked for includes directions from transit stop to the store entrance, landmarks that can identify the store, etc. The users reported a significant ease in navigation using GPS devices when walking on mainstream streets as opposed to walking inside an OSC. All our participants reported that

seeking help from nearby humans would be their final resort. A summary of the key findings in the case study is given below:

- An on-demand description of their surroundings is always helpful for them to orient themselves.
- Description of the surroundings can be effectively done in terms of egocentric or allocentric methods. The usability and preference of this description varies widely from user to user and by location.
- Some users prefer to have the description of distances given in steps, as opposed to feet; low-vision users may still enjoy the availability of a zoomed map.
- Tactile landmarks are preferred to assert the location.
- Using angles for direction is not usable. But the users are familiar with terms like “diagonally right/left”.
- It is not a good idea to direct them to walk through parking lots. It is often unsafe to do so and would involve hitting the cars with the cane.
- Longer safer routes are preferred over shorter routes with parking lots or obstacles.
- Extra information: Restroom location, traffic conditions on streets encountered, user created markers for future reference, etc. would be great add-ons.

Part of the study involved asking our participants to navigate a chosen local shopping complex with the help of the two aforementioned iPhone apps. It is to be noted that these apps helped only for navigating the major streets bordering the shopping complex but failed to provide much assistance for navigating the complex itself.

3.3 Design Guidelines for Building Apps for OSC

Based on our case study, we established a set of guidelines for developing an effective iPhone app in addressing the deficits of existing solutions. We summarize them below:

Avoid additional screens as much as possible: A VI user would face frustration in using finger gestures to get back and forth in a multi-layer menu and thus a flat structure should be used as much as possible in the interface.

Less is more: Our users disliked navigating through a page filled with too many buttons. They asked for a few functionalities that convey a lot of information, during our case study questionnaire.

Layered information delivery: Users with varying abilities will need different amount of information. Having information interlaced with gestures is preferred, so additional information is presented only on-demand.

Orientation and Mobility training: Most of the blind users have undergone the O&M training. The app needs to be consistent with the protocols in presenting the information.

Supporting user notes: Every user from our case study picked up some different cues about his/her surroundings. A mechanism for the user to record his/her own markings would enhance the usefulness of the application.

Supporting user customization: Low-vision and completely-blind users have different requirements. Users may prefer measuring distance in different ways. The app should allow some user-level customization to support such features.

We now present the design of a novel iPhone app that aims at addressing the challenges faced by VI users in navigating OSC. Largely based upon the design

principles which were derived from the exploratory study, the proposed design and implementation attempts to overcome the challenges from the following four aspects: an information fusion technique, an information representation structure that is appropriate for mapping-related tasks in an OSC setting, an intuitive interface and interaction scheme, and the support for user-customization to cater to individual needs of the users. These are shown in the figure 13.

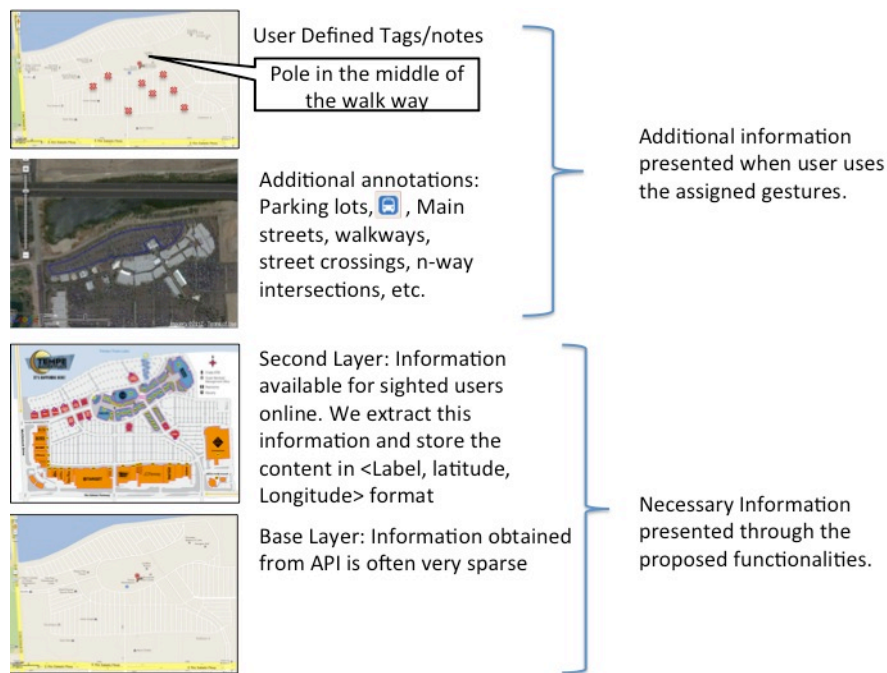


Figure 13 : The tiered information representation scheme used to support application.

3.4 Tiered Information Creation

The major part of the problems faced when using map-based technologies for navigation in OSC is the lack of required information. Publicly available map services such as Google maps do not have the desired level of details for typical OSC. However, most shopping centers maintain and publish maps with rich annotations.

Also, a sighted volunteer may be able to label a satellite image of a shopping complex as to where are the parking lots etc.

Finally, a user while using the app, may want to insert his/her notes to a location. Considering all these possibilities, we adopt a tiered representation for all the mapping information. Figure 13 illustrates how this is currently implemented. In Figure 14, the base layer corresponds to the map that is typically available from a GPS system, the second layer is the layout map given by the shopping center, the third layer illustrates an image of the same locale with additional labels, and the fourth layer is used to store user notes.

3.4.1 Base Layer Information

As seen in Figure 13, the base layer contains information available in the typical mobile applications using Google/Bing/Apple/other maps. This information is sometimes sparse, and from our studies, often inaccurate in an OSC setting. The locations of individual stores and bus stations are more often than not, inaccurate and haphazardly placed. Using this information for VI users is often dangerous for this reason. In our scheme, we find additional sources to obtain this kind of information with reasonable accuracy and map it on top of the base layer.

3.4.2 Second Layer: Information from the Web

The most vital information when navigating an OSC is the location of the stores inside the mall. This is the kind of information that is not available on the Google maps. One way to obtain this sort of information is to devise a method to integrate the Google maps with the store directories available on the website of OSC. . This can be viewed as a map-to-store map registration problem. We try to register the

shopping mall directory to its corresponding Google map. The purpose is two-fold. Shopping directory has much more information as compared to maps and by overlaying shopping directory onto Google maps, we can still retain the GPS coordinate information.

Control point detection is a major step in any registration problem. We propose the use of shop centers as the control points as the shapes of stores in both Google maps and shopping directory have high-level similarity. We also employ road and parking lot detection but are not used for extracting control points since shopping directory may not have them. Firstly, we detect yellow-colored major roads and orange-colored freeways from Google maps by simple color segmentation. Black-colored text labels on Google maps are detected by using the same principle. Roads in Google Maps always have text labels on them and they have a lower-limit on their width. We use this fact to distinguish parking lots from roads. The roads are detected by using region-growing image segmentation technique in combination with distance transform. Distance transform allows us to monitor the width of the road. If it falls below the predefined minimum width or if there is no “white” road, then the region-growing stops. Text labels act as seed points while performing region-growing.

3.4.3 Supporting User Customization

The necessity of supporting user customization has been concluded earlier. In our current implementation, we support the following three types of customization. The first customization provided is the user preference on distance metric. Some users preferred the usage of steps and blocks to usage of absolute distance in terms

of feet. We introduce a user-based calibration feature that calculates the step-to-feet ratio. The user is asked to walk 20 steps prior to the visit in a familiar environment to allow for calibration. The second customization is the method used to convey direction. Users prefer egocentric or allocentric method of description, depending on the location and complexity of the surroundings. The third customization provided is based on the level of vision of the user. The totally-blind users interact with the system using the predestinated gestures and through voice-over output. The low-vision users can have the additional freedom of interacting with a spatial zoom-able map and larger font sizes.

3.4.4 Interface Design

Based on our case study, user input and our understanding of the existing GPS devices, we propose the following interface to help users navigating OSC. Our application uses iPhones accessibility mode and uses the standard voice-over gestures. A user can scroll through the buttons without knowing their spatial locations, using one finger swipe. A double tap anywhere on the screen selects an item. We refrain from introducing too many additional gestures to make the app as simple to use as possible.

Homepage: The homepage of the app consists of an entry into all the possible functions of the app. We adhered to the design paradigm inferred from the user study, stating the user preference of not having too many screens to navigate and having a few necessary functions on the screen. We have designed our homepage in such a way that the user can obtain all the important information by staying on the

homepage. We interlaced various gestures to help access additional information. Figure 2 shows a screenshot of our homepage and an overview of all the functions.

Where am I? “Where am I” is a popular functionality provided by most of the existing GPS applications. A sample result for this function can be illustrated as follows: “Facing North near 661 Meadow Avenue”. Once inside a shopping space, a description in terms of an address is no longer relevant information. We modify this functionality to make it more informative and convey necessary information in a tiered manner. When a user double taps on the “Where am I?” button, our application reads out the orientation of the user, the closest landmark and the nearest landmarks in either directions as shown in Figure 14. He repeats this gesture to listen to this information again. The user has an option to ask for more information after he listens to this information. A pre-assigned gesture (3 taps on the iPhone) provides information about the nearest streets and any user tagged notes if they exist around this location. He repeats this gesture to listen to this information again.



Figure 14 : An overview of iExplore's homepage and a sample output for each of the proposed functionality

Points of Interest: Most applications and GPS devices support points of interest, where a predestined number of landmarks around the current location are listed out to the user in terms of distance and direction. We include this feature in our application, except that the information associated with this button is according to the tiered representation discussed above.

Where is my destination?: The directions provided by most applications are hard to use for VI users. We propose a scheme to provide blind-friendly directions. Once inside an OSC, the user can be standing inside a parking lot, a store or at a major landmark. Our description of the destination location takes into account that users do not like walking through the parking lots, considers the safety quotient of streets, includes description of places in-between in terms of stores lining the route and major streets on the way. The user inputs the name of the store or landmark he is interested in using the speech input feature provided. A verbal description (as seen

in Figure 14) is then generated tailored to the user preference of egocentric or allocentric methodology, in terms of their distance preference. We use Dijkstra's algorithm to compute the shortest path among the walkable options. Using this path, our description takes into account relative positions of the landmarks and additional tagged information about the surroundings.

3.5 Experiments and Analysis

After registering Google maps to the corresponding shopping directory, we develop an evaluation scheme to validate proposed method. It is described as follows: We detect stores in Google maps as well as in the registered image. Now, we estimate the percentage of pixels of the Google map which overlap with the pixels of registered image. This also validates our shop detection algorithm. Figure 18 show the result of shop detection onto the Google map and the shopping directory. There is a 66.45% overlap in the area marked by our approach when compared to the original store directory and the ground truth. The parking lot detection is 100 percent accurate.

3.5.1 User Feedback on Overall System Design

Upon loading the app on the user's iPhone, we asked them to take time and get familiar with the buttons and tabs on the home screen. The users were asked to think aloud (make explicit comments) while using the app. They were not provided any additional training on the usage. After the users were familiar with all the functionalities, they were asked questions about the interface, its usability, the intuitiveness of each button, easiness of navigating the buttons, etc. The users were encouraged to ask us questions to clarify any aspect they found confusing. One key

observation every participant made was that the interface has very few selectable items on the screen. They were able to summarize the key components upon closing the app. They also reported to have received more useful information than they would have hoped for (compared with other systems) from each functionally buttons.

3.5.2 Testing the Proposed Functionalities

The users were taken back to the testing site, where we conducted our initial case studies. We started at the bus station for uniformity and to simulate a real-time scenario. The users were asked to test each of the functionalities and were asked to walk around. We tagged them with a volunteer for safety reasons and the volunteer collected feedback from the users. The user was asked to change the settings to his preference of distance, terminology for directions, etc. Once done, the user is asked to go back to the home screen. He is asked to find the “Where am I” button and access the information. We then asked the user to point out the direction using their hand, towards one of the landmarks mentioned. This test was aimed at testing out the accuracy of information conveyed as well as user understanding from the verbal descriptions provided. The users got their directions right with 100% accuracy.

3.5.3 Evaluating Navigation Support

To evaluate the support provided by iExplore for navigation, we defined preset routes with 2 turns and asked the users to use the app to find the destination. If the participant took a wrong turn, we recorded a miss for that segment of the journey and let the user turn around. The user was asked to use the app to gain a better understanding of his surroundings. The system failed to orient the user in the right direction once and one out of the four users asked for assistance once during

the experiment. At the end of the stretch, the user was asked to describe his surroundings and the relative positions of the store he walked by. This test aimed at validating our assumption that the users liked exploring and gaining information about their surroundings. We did not intend this test to be a memory game, but most users were able to figure out the relative positions of the stores they walked by with 100% accuracy.

3.6 Sustainability

After the experiments, the users asked for a version of the app that they can download and use. While this was encouraging, we faced the problem of aggregating the required data to support a big space. We checked for data accuracies on the routes that we used for our experiments. We couldn't possibly do this exercise for every route in the mall. We had volunteers collecting GPS coordinates of every major and minor landmark manually. We had trained volunteers who knew how to identify the details that are important for individuals with VI. For this system to be sustainable and deployable, we had to find a way to obtain this information and have quality control over the accuracy. We then started looking into methods to check for accuracy our data from our image processing techniques. We started exploring the idea of using crowdsourcing for getting details for maps. We wanted to present strangers with results from our methods along with an interface that helps them indicate landmarks and rectify any mistakes made by our detection methods. It employs a combination of crowdsourcing, human-in-the-loop and image processing techniques to give detailed and customizable route between two places. Though our

focus is on the navigation aspect of the app, we briefly describe all of its features, other useful innovations and its interface for completeness.

3.6.1 Front End Interface

The design of the user-interface is largely based on our observations obtained through our subjective study. The interface design is also influenced by our study of existing GPS devices and navigational apps for VI users. The interface is completely accessible and uses standard gestures. It also integrates seamlessly with the iPhone's default voice-over mechanism. The detailed features of our app are described as follows.

3.6.1.1 Home Screen

The home screen is designed by observing the operating style of VI users. They mainly use horizontal swipes to navigate through all the screen buttons instead of locating them by guessing their positions. The button selection starts either from the top-left corner of the screen or from the last contact point on the screen. It continues laterally in the direction of the swipe. Therefore, for VI users, quickly (and easily) accessible button positions lie along the top edge and directly above the *Home* button on the iPhone. Hence we have placed commonly used buttons along these positions. In the bottom row of buttons, we include another button, called "More/Less", to control the amount of information being presented to the VI user. We use a layered representation to regulate the flow of information according to user's preferences. In the middle of the screen, we include a map. VI users can use it to explore streets and places around them. Sighted people can interact with the map to better guide VI users. The home screen of our app is shown

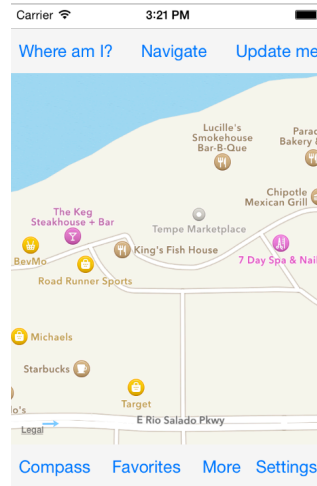


Figure 15 : Home Screen of our App.

on the left in Fig. 15. We will describe the layered information representation and the function of the map in a later section. Details of individual buttons are as follows.

Where am I? : This is one of the most popular features in existing navigational apps. It is supposed to inform the VI user about the surroundings. Many existing navigational apps (for example, Ariadne GPS) tell the street address or let users explore the places by touching on the map. There are drawbacks with both options, as VI users cannot interact with the full-screen map precisely, especially when there is no reference point present on the screen. Street address is almost always not useful in an OSC setting. Thus we introduce layered information representation. In the first layer, VI users will only here street address. In the second layer, they will hear the names of nearest shops in all four cardinal directions along with the layer one information. Such kind of representation provides extra information to a VI user on-demand without feeding excessive information. VI users can select the layer by pressing the button “More”.

Compass : This button informs the user about his heading. The description of the heading is available in three ways: four cardinal and four ordinal directions (e.g. North, North-west etc.), degrees and clock-system with respect to user's current heading (for example, 180 degrees or 6 o'clock indicates a direction behind user).

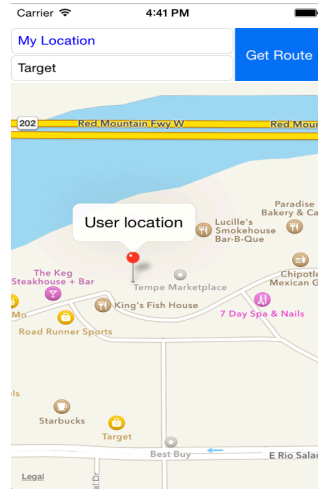


Figure 16 : The user side of Navigate.

Favorites : Many VI users prefer to store certain points depending on their requirement. Some of the typical examples include shop entrance, restroom location, water fountains etc. However, existing apps only allow a text label for such entries. Through our study, we figured out that typing on a smartphone keyboard is an unpleasant experience. To that end, we introduce voice-labels to allow VI users to be able to input long descriptions if they need to. Once they press the “favorite” button, they are asked to speak their description of the place into the microphone. When they pass by that place in the future, they will get a vibration on the device. By double tapping, they can hear their description along with the information about the surroundings, which is automatically inferred from user's location. We call this functionality as voice-labels. Vibration can be switched off from the Settings so that

every time the description will be heard. VI users can also disable the voice-label feature and choose to type the description.

Settings : It includes all of the settings described above in the functioning of each button. Along with that, basic app settings are also present. For example, a VI user can change default distance unit from foot to meter or yard. The user can also set the default layer representation when the app is launched.

Navigate : As mentioned before, a novel navigational scheme is the major contribution of our paper. Crowdsourcing, human-in-the-loop and image processing techniques are applied together to provide a rich, accurate and safer set of directions to the VI user. We term it as *Assistive Navigation*. The central idea of this scheme is as follows.

A VI user seeking directions from his/her current location to any place within the OSC submits a request through the app. The request is submitted to our server. We have a set of sighted users who have registered on our website. One of them chooses to provide updated directions for the incoming request. The set of updated directions is usually given by looking at the satellite view using a mapping service of their choice. As soon as the set of updated directions are updated to the server, the VI user will get a notification. The VI user can follow the directions to get to a shop. The shop's position is as indicated by the map service. Most of the widely used map services will point the user to the center of the shop. However, the span of a large shop could go up to 500 feet. Hence, finding the entrance after having been traveled to the center is a tiring experience. We use simple-but-effective image processing techniques to correctly determine shop-spans. Therefore, our system notifies the VI user as soon as he/she comes in the vicinity of a shop.

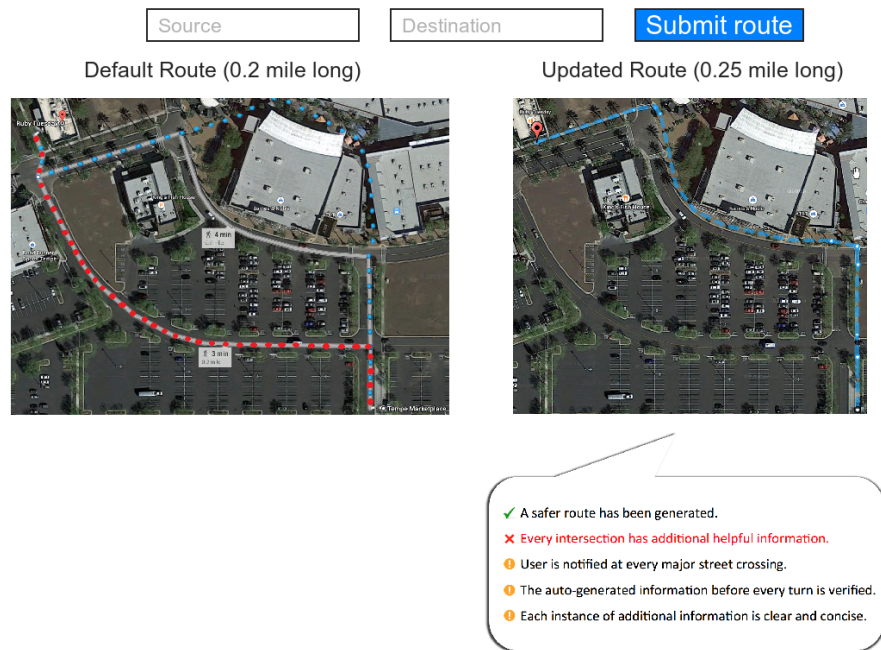


Figure 17 : Server-side interface with mandatory checklist. Volunteers see this after accepting a request from a VI user.

In the following subsections, we dive into details of each component of this scheme.

1. VI user-side interface: The interface pertaining to this scheme opens in a different screen after he presses the “Navigate” button on the Home screen. The interface consists of two text fields where the VI user has to input his/her destination and the starting point (if its not the current location). There exists a button called “Get Route” which submits the request to server. Unless the VI user’s current location is the starting point, an accurate address is required; otherwise system will reject the request. For places in an OSC, it is difficult to obtain an exact address of a place. Therefore, we provide a map view, where sighted users can touch on the map to record a starting point in terms on latitude-longitude. Interpretation of destination is usually unambiguous since it exists near

the starting point. We used Google Places API to get the address of places nearby just from their names. The interface is shown in Fig. 17.

2. Server-side mechanism: As mentioned before, a request from a VI user gets uploaded on the server, which is then processed by registered users (called *volunteers* henceforth). They see each request in the form of *source to destination* along with the travelling distance. Once they select a request for further processing, they see the entire list of paths generated by Google maps. Once they select a relatively safer route from that list, they are asked to follow a checklist before they can submit the updated route. Fig. 17 shows the entire process with the mandatory checklist. The list of all the paths from *source to destination* is generated with the help of Google maps API. To guide volunteers select a safer route, we prepared a list of instructions for them to follow. This list is given to them at the time of registration. The instructions are based on our study of the structure of Google map, existing GPS and navigational systems and of course the needs of VI users. The main instructions are summarized as follows.

- Always choose a safer route even if it is a little longer.
- Satellite view can show a sufficiently detailed view to enable you to mark points on a sidewalk or a pathway, instead of marking on road.
- You will encounter three kinds of roads in Google map. A thick white road indicates a major Street whereas thin white road indicates a local street. A gray road indicates a walking trail. Be sure to check both satellite and map view.
- Be sure to notify the VI user at every intersection and every street crossing. The major street crossings should be mentioned explicitly.

- We automatically generate instructions before every turn and street crossing (after you insert it) to notify VI users. Make sure to verify those instructions.
- Make each piece of information clear and concise.
- Information about distance and heading between two points is auto-generated and should not be changed. Your additions should come after these instructions and they should maintain the flow between the sentences.

The volunteer can drag points around to modify the route so that it adheres to the above guidelines. We also provide graphical illustrations and examples for each point. To ensure a high-quality route, we employ error checking in terms of *proactive prompts*. These prompts are shown to the volunteer just before the route is submitted. They are concise prompts that remind the volunteer about the necessary steps. Green checkmark sign means that constraint is satisfied.

Volunteer has to click on orange constraints to indicate that the appropriate care has been taken to satisfy them. Red constraint means that our system has detected an error in the updated route, which will prohibit you from proceeding further. We will talk about error checking in the later section, which describes image-processing techniques. Once all the prompts turn green, volunteer can submit the route.

As mentioned above, volunteer only needs to drag points around to modify the route. Since all the instructions are auto-generated, volunteer only needs to add one to two sentences at few points such as intersections. Thus the process is fast. The updated route is verified through proactive prompts and error checking.

3. Presenting directions on the VI user's device: The directions are presented to the VI user as he/she approaches the checkpoints marked by the volunteer. The VI

user can repeat any instruction by double tapping with two fingers on the screen. Please note that while repeating instructions, the distance and heading information will change accordingly. We notify the VI user 10 meters before the actual checkpoint. This is done to compensate for the 5-meter error margin of a normal GPS. Also, the voice-over requires some time to speak the instruction. We expect the VI users to find the street crossing or a shop in the vicinity of 5-10 meters using their white canes. This has worked well in practice.

4. Image-processing-based techniques to determine shop vicinity and to perform error checking: Most of the mapping services show the shop locations in the center of the building. For large shops, the difference between the actual entrance and the center could be as large as 200 feet. To address this issue, we use the static Google map API to obtain the map that is centered at the shop location and with a specified zoom level. The next step is to perform region filling starting from the center of the image. The key observation here is that a shop has same color in the static map. In this map, minimal amount shop labels are visible. Region filling stops when it encounters the shop borders since the color changes at that point. Simple post-processing such as filling holes produces a mask as shown in Fig. 18. The red circles shown in Fig. 18 are corners detected by a standard Harris corner detector. When the VI user's distance to one of these corners is less than a preset threshold, an alert is produced.

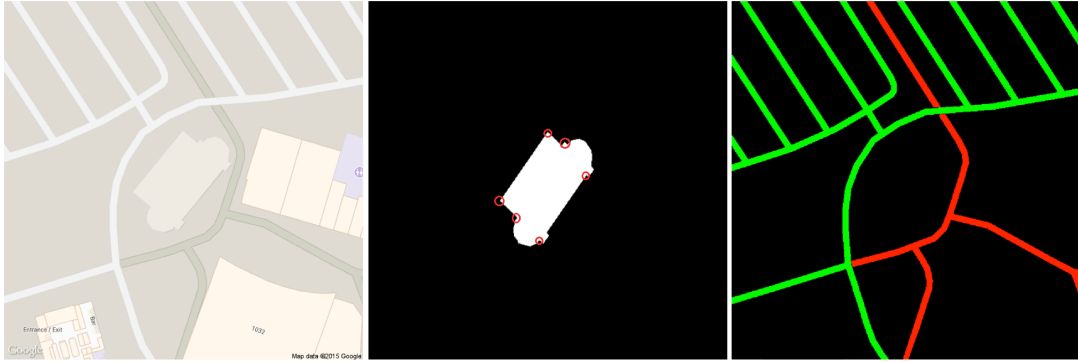


Figure 18 From left to right. Static map view centered at the shop location. Detected shop span. Red circles represent detected corners. Detection of different types of roads from Google maps.

Error checking happens when the volunteer is about to submit a route. This helps to filter out any inadvertent errors from volunteer's side. Our key observation is that roads in Google maps have specific colors. The roads are detected by simple color thresholding and connected component analysis. The detected roads are also shown in Fig. 18. Green and red colored road indicate streets and walking trails respectively. For error checking purposes, we are only interested in streets and all intersections. We check for two types of errors. We warn the volunteer if majority of the route lays *directly* on the green colored road i.e. a street instead of a walking trail. We detect intersections of all kinds and make it mandatory for the volunteer to add some useful information (e.g. street crossing ahead).

Update Me : This button updates the instructions to the next checkpoint with respect to the current position of the user. It is equivalent to double tapping on the screen with two fingers as mentioned before. The direction and heading of the VI users will change dynamically as they keep walking.

3.6.2 Evaluation

them on using the phone as a part of the experiment. All of them have been using the iPhone for more than two years. All the users are relatively tech-savvy and mentioned that they use other navigation applications for getting around in their everyday life.

Location : Although our techniques support the navigation on any OSC, we picked a local mall for the same of convenience in testing. We tested the users ability to get to 3 destinations from 3 sources. The 3 routes were chosen based on the lengths and the complexity. The users had to input the destination they wanted to visit and wait for the directions from the application. The shortest route involved walking along the sideway, crossing a street and walking some more. The longest route was about half a mile long, featuring curved pathways and major street intersections.

Server Side Interaction : One of the designated volunteer would be scheduled to be online during each session of our experiments. Upon getting a request, he would use our server side interface to do the appropriate markings. This information is used to generate navigation information by the system and sent back to the user. The volunteer's job is to look at the suggested route and edit it out based on the prompts our system gives him. They rated the interface a 5, on a scale of 1 to 5, with 5 being easy to use and 1 being frustrating. Every volunteer was able to finish his assigned task in less than 2 minutes.

Our experiments consisted of two phases. In order to give the users a reference what it would be like to navigate without our application, we start off our study by asking the user to navigate to a store of our choice using any application of his choice. Two of the users requested to skip this exercise because they found it too

stressful. Two of the users called the store and asked for directions multiple times. One user asked the passersby for directions.

The second phase of the experiment consisted of the users requesting directions to the predetermined stores of our choice. They were instructed to follow the directions and we recorded their travel. The users were told that they were allowed to quit a route at any point in time if they feel unsafe or find it stressful. After each destination is reached, the user is asked a fixed set of questions. We recorded the difficulties faced and any ambiguities that they point out. They were asked to rate the overall usage experience, quality of directions provided and amount of information provided.

3.6.3 User Feedback

Evaluating directions based on users performance on a route is tricky since navigation is a combination of proper directions, users mobility skills and the state of the environment. We faced varying conditions like peak traffic, extremely noisiness at the mall and users who are nervous about travelling to an unknown destination for the first time. We are happy to report that we did not have incidents where the users quit a route due to its complexity or from being frustrated with our system. Two of the users exclaimed about how they are taken back about the accuracy of the directions. The app scored an average of 8.2 on the overall experience. The average score on quality of directions is a 7.6. The average score on amount of information presented is an 8. But this feature was customizable where the users can indicate how much information they want to hear. Our recommended setting was level 2 (midway), since none of them used this app before to know the

amount of information that can be delivered. We got a rating of 9 on ease of accessing the information. The wait time was rated an 8. But the users had remarks on it like “I don’t mind since there is nothing else like this” and “Can I queue in requests before I reach the location”. Most users indicated that they were highly likely to use it and recommend it to their friends.

3.7 Conclusion

Safety is a major concern when developing an automated system that guides an individual with visual impairment. There are many systems out there that could not be deployed beyond research due to similar concerns. We created a workflow for a previous work of ours that intends to automate the process of helping users navigate an OSC. And enlisted the help of crowdsourcing to validate our computer vision techniques. The experiments support our claim that a reinforcement system greatly benefits the backend methods when the application involves users who are visually impaired.

CHAPTER 4

A VIRTUAL ORIENTATION & MOBILITY INSTRUCTOR

4.1 Problem Statement

Twenty million students reportedly will attend university in 2015 [21]. Being able to adapt to the campus life is critical to freshman retention, and thus a wide range of orientation programs and resources are available on college campuses. For a typical student, finding classrooms and commuting between buildings for classes, libraries and cafeteria are part of nearly-effortless daily routines. Unfortunately, for individuals with visual impairment, such tasks make the acclimation process very stressful. Unlike street navigation that may be assisted by popular mapping services [22] [5], a typical campus differs in many ways: there may be numerous pedestrian-only streets in close proximity; there may be many no-name walkways; there may be many buildings in a small space that a student needs to know about; buildings may have multiple entrances/exits on different sides, etc. Traditional mapping services [23] would fail to provide adequate information for a visually-impaired student due to such complicating factors.

Many incoming students with visual impairment seek to get familiar with the campus through formal training with Orientation and Mobility (O&M) instructors and/or informal training with family/friends. O&M training is a one-on-one exercise, during which an instructor walks through the campus with a user and informs the user all useful orientation and navigation information. This is a time-consuming exercise that may also be costly, and hence many students may seek less formal training with family and friends. One limitation of such human-based training is that it is not possible to cover every route. Also, while a training session may cover a

lot of information, some may not be used until days or even months later and thus may get forgotten before they even get used. Consequently, the unfortunate reality is that, even with some O&M training, most visually-impaired new students would still feel it is stressful to get around on campus after a long while.

Recognizing the (practically unattainable) benefit of having an O&M trainer anytime and anywhere, and also inspired by a specific blind student who struggled with navigating the campus for the first six months at the university before he decided to hire an O&M instructor to teach him the routes. In this work we propose to design and develop a smart phone application that attempts to provide some basic functionalities of such one-on-one training. With such a “virtual trainer”, a student can learn a route at their own pace whenever the need arises without having to rely on a human trainer and without having to worry about remembering everything from a single training session.

The key idea of the proposed approach is to employ the accessibility features on iPhone coupled with data mined from Google Map and crowdsourcing to develop a virtual O&M training app for a given campus. The principles for O&M training as established in [24] are used as guidelines in our design for developing a schema for training the user. Factors like orientation, ability to understand scale, movement and ability to use other senses are essential for successful navigation. We do not presume to teach these basics but instead focus on training the users to learn the new environment. We examined the current practices used by O&M trainers and attempt to mimic some of these methods in our application.

There have been some related prior efforts catering to developing solutions [25], [26], [27], [28], [29] specific to such spaces for individuals with visual

impairment, each addressing certain needs such as establishing the current location or giving a route. While such needs are also considered in our work, the focus is on building a virtual O&M trainer rather than a navigation assistant as done in most prior efforts. Our study suggests that a virtual trainer serves our participants very well and a full development of the idea can potentially fill a critical void in the literature.

4.2 Proposed Approach

This section outlines our approaches to realizing the objectives discussed in the previous section via a smart phone app. We start with an overview and follow up with discussion on key modules of the proposed application. An O&M instructor would typically walk along a path from one landmark to another while training a user around the path. Our app follows such a convention by asking a user to define a path. In a nutshell, the app is designed to work as follows. When a user launches the app, he enters the home screen. He may then choose either “overall description” (or offline training) or “onsite training” (or other non-training buttons detailed below). If he chooses the former, he will be prompted to enter two landmarks (via voiceover) and then he will get an overall description on the walking path between the two landmarks. If he choose the latter (which means he is in the on-site training mode), he will be prompted to enter a destination and then a path from his current location to the destination will be defined. Then he will first get an overall description of the path and then segment-by-segment training as he proceeds along the path.

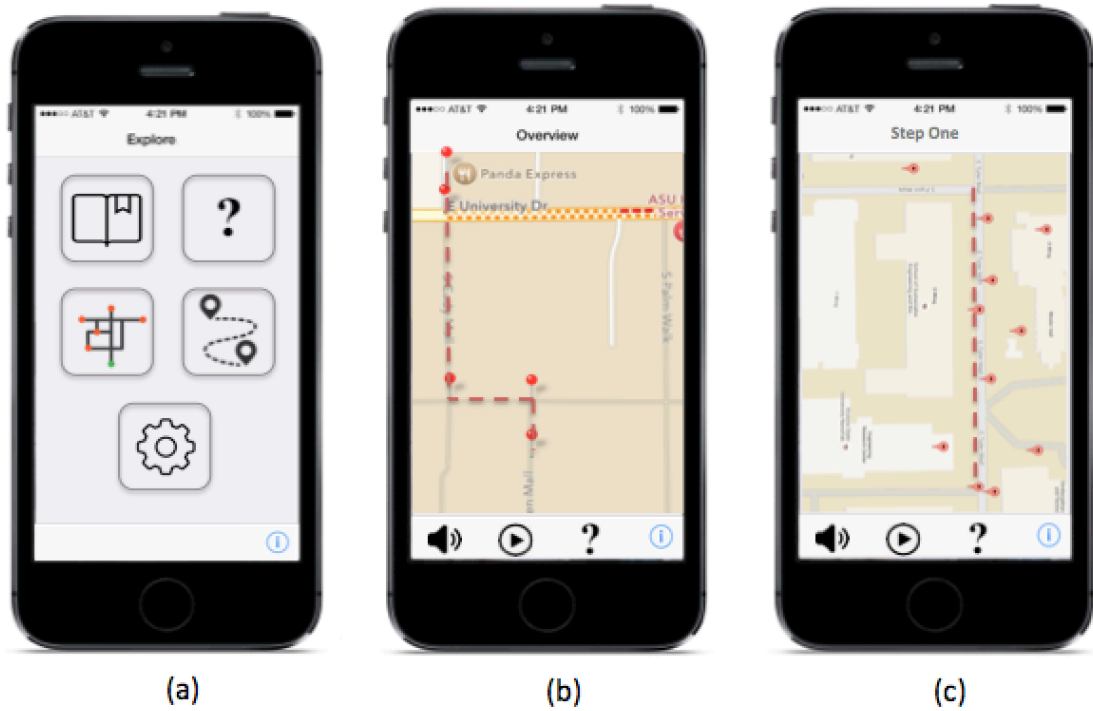
Our app organizes and provides information in tiers. The first tier contains the street and building names. This is what the user would get as an overview. The

second tier is additional large landmarks that help learn the space. The third tier contains elaborate information such as entrances, water fountains, etc., which is delivered on demand based on the position of the user as recommended by the literature [30]. O&M training differs from navigation assistance as the goal is not to get a user from Point A to Point B, but to teach about the space, and thus the process requires considerable time and effort from the user. For effective training, we break a route into segments and organize information around the segments. The app asks quizzes during training to confirm a user's understanding of the space. The user needs to pass the quizzes to go on to the next segment of the path. The proposed app requires a finer level of details of the landmarks that Google maps do not typically support, such as entrances for each building, textures of walkways, restrooms, water fountains, vending machines, unmarked walkways, etc. We relied on community sourcing for gathering such information for the specific campus in our study. Given the ubiquity of smart phones, such a strategy may be extended to other campuses if a proper server is built to accept inputs from volunteers.

4.2.1 Interface Design

We use the wisdom of the popular applications in the market (<http://www.ariadnegps.eu/>) to design the interfaces. as illustrated in Fig. 20. Fig. 20 (a) is the home screen, which lists five functionalities: bookmarks for locations/paths, “Where am I?”, new route training, overall description of a path, and settings. “Bookmarks” are for storing shortcuts to favorite paths/locations, while “settings” collect user preferences on how they like their directions delivered and amount of

haptic feedback expected. The other three functions are elaborate below. For the benefit of low vision users, the screen also depicts a map during some steps.



Step 1: You start on South Palm Walk. Proceed North on South Palm Walk for 261 feet. Then, take a right on East Tyler Mall.

Step 2: You reach the intersection of South Palm Walk and Tyler Mall. Proceed East on Tyler Mall for 778 feet. Then, take a right on South McAllister Mall. The following two intersections are in your path. Tyler Mall and Farmer Mall. Tyler Mall and College Mall.

Step 3: You are on the intersection of Tyler Mall and South McAllister Mall. Proceed South on South McAllister Mall for 405 feet. Then, the destination is on your left.

Figure 20 : An overview of our application.

4.2.2 Overall Description / Offline Training

Under this, a user obtains relevant information along a path between two points before traveling the path (or even before coming to the location). A user may use this mode for (i) accessing details of a route as part of offline training before visiting the site, or (ii) getting a quick overview right before getting trained on a path (the on-site training module starts with this overall description). We utilize

Apple maps API coupled with the data we collected to generate the description. A sample is given in Fig. 21.

4.2.3 Where Am I (WAI)?

This allows a user to query for his location. This is available from the home screen and any step during on-site training (e.g., Fig. 21). The presence of a large number of landmarks around makes presenting them rather challenging. We use a combination of cardinal directions or relative positioning coupled with the clock system to generate an answer.

“You are on South Palm Walk facing North. There are not landmarks between your 3 o clock and 6 o clock. There is the bookstore in 96 feet at your 7o clock. In 150 feet, there is Café Java city at 9o clock. In 200 feet, there is Engineering Research Center at your 11o clock.”

Figure 21 : A sample where am I.

Input = [Start Location : End Location]

1. Verbally describe the overall route from Start to End location.
2. Familiarize himself with the landmark.
3. Point to objects from the landmark.
4. Travel to those objects from landmark
5. Point back at the landmark from the objects
6. Travel back to the landmark from each object.
7. Point to the landmark from specific objects in the environment with known relationships to the landmark, having not started at the landmark.
8. Travel between objects with known relationships to the landmark w/o returning to the landmark.
9. Return to the landmark. Proceed on route.
10. Repeat this process every few yards / major landmarks till the end location is reached.

Figure 22 : An overview of the steps involved in training.

4.2.4 On-site Training Mode

This is when a user walks along a path while getting trained. In developing our training routine, we utilized the standardized O&M manual [24], which places

an emphasis on the landmarks, the clues, the outdoor numbering system, the measurement and the compass direction. The major steps include: familiarizing with the start point, detecting surrounding landmarks, observing traffic conditions, travelling between current landmark and surrounding locations, figuring out route to travel, and answer questions asked by the trainer about the information learnt. We aggregate these steps into a training process outlined in Fig. 22. Each route starts with an overview. Then we present with the information about the surroundings. The user may be asked to travel to one of the landmarks that is close by and back to the start point. He is then asked to proceed on the route using information that we provide. The app determines the next position to stop the user and repeat this process based on distance and number of new landmarks along the way. The user is asked to take a quiz every few hundred yards to make sure he pays attention to all the information presented. In human-based training, this step is administered by the O&M instructors to test the retention of what has been learnt. We mimic this by testing a subject's understanding of the directions. The app asks the user a set of questions at each step. Quiz questions are auto-generated using the landmarks encountered and the known relative positions. Sample questions are like: (1) "Is Engineering Research Center closer to you than the Physical Sciences building?"; (2) "Point the phone to the direction of the Science library". User answers are graded automatically on the spot before the training continues. Figure 23 illustrates a sample route and the implementation of the training. As the training app does not intend to give only directions but also comprehensive training mimicking a guide walking with the subject, the process can be longer than the

typical time required for walking the path (see Experiments). This is in consistent with human-based training.

4.3 Experiments

The application we designed is intended to mimic a human instructor training an individual with visual impairment. We do not expect the training module to do everything a human trainer can. We designed a study [17] to test and validate the basic functionalities of the proposed app. We picked two equidistant routes on a local campus with similar layouts. The routes were half a mile each, had the exact number of turns (two) and similar traffic density. The first route was used to study how the participants travelled without our assistance. The second route was picked to let our users test our application. Each of the routes had three segments. One segment is fairly well known on campus, one not so well known and one very discrete. This ensured varying levels of student crowd during the experiment.

4.3.1 Participants

We conducted the experiments in two rounds. The first round involved us inviting two Orientation and Mobility instructors to come and test out the application. We wanted their nod before we introduced it to our target audience. We recruited six participants with varying levels of familiarity with the campus of Arizona State University. Two of the users are congenitally blind. Three of them are cane users and the rest use guide dogs. They rated themselves as fairly active people who travel to unfamiliar places on regular basis. All of them are iPhone users who are familiar with existing applications like BlindSquare (blindsquare.com) and Sendero GPS (www.senderogroup.com).

4.3.2 Feedback from O&M Instructors

We had two O&M instructors come in and test our application. We introduced them to two routes that we wanted the app users to test out. They were asked to give a generic version of the O&M training to one of the developers on the first route. We recorded their protocol to learn about their training style. They were then asked to use our application and give their feedback on the second route. We realized that O&M training is a very personalized process. Both the trainers agreed that attempting a generic training protocol is a very hard process. It's very dependent on the users cane skills, hearing skills, cognitive maturity and his understanding of time-distance. The standard protocol requires them to assess their audience before trying to teach the student a route. We explained it to them that we are targeting the age group 18 years or older. If they did the training as required by the State government, they should have the cognitive abilities to use our application.

4.3.2.1 O&M Instructor One

One of the instructors (aged 36) is visually impaired and gave us some unique perspective. She usually trains young kids as her day job. She trains adults on a need basis. When we asked her to evaluate our system, she asked for the details of the routes we would be using to test the application. She split up her task into two. We provided her with a campus map that she could access using a magnifier and gave her a list of the routes we would be testing the application on. She went to the location ahead of us meeting her and she explored the path by herself. She has partial sight and hence wanted the extra time. She explained that this extra step is a limitation associated with her being visually impaired but she can really relate to

the students she trains. When we first met her and explained to her our intention of developing a training app, her initial remark was: “it is not possible to create an O&M app.” We asked her to navigate the first route for our sake to learn about her style of training. She underwent the training as recommended by our app on the second route. She gave us her rating in terms of the following five important aspects of training:

1. Time-distance understanding: Rating: 4/5

Comments: “It is very important for the user to understand the distance covered with reference to time. That is what helps the user to stop after say 100 feet. I like that this app gives you distance covered and distance to the next intersection.”

2. Accuracy of information: Rating: 5/5

Comments: “I like how you have information about non visual landmarks like the fountain and coffee shop.”

3. Usefulness of information: Rating: 3/5

Comments: “You have gathered a lot of information. Some folks need all of this information. Some don’t. But there is no real way to assess this in an app. A human instructor has this advantage. But I like the quizzes and tasks.”

4. Availability of rectification information: Rating: 4/5

Comments: “I like that I can pull up the compass to rectify any veering. I like that if I miss a turn, the app tells me. It vibrates when I am close to a turn.”

5. Quality of training: Rating: 3.5/5

Comments: “I am surprised by what this app could achieve. I like the amount of information it provides and the activities you make the users do. The only downside

I see is the lack of customization with amount of information one hears. To a few users, it may get overwhelming and they shut down. But this is a good app.”

4.3.2.2 O&M Instructor Two

The second instructor (aged 52) is a sighted individual who has been training individuals of every age group for the last 30 years. He has trained more than 30 people on the campus of XYZ University before and is very familiar with the space. On route one, he points to many subtle clues that aide navigation that we didn’t know of. For example, there was a subtle up incline on a walkway, which flattens at an intersection. When collecting landmarks/clues, we missed a lot of such details. He blindfolded one of the authors, provided her with a cane and helped her navigate the route. He detailed the route in excruciating detail. Every tree, bike racks, texture of the floor, canopies, buildings, etc. were remarked upon. He made the author listen for clues near every building and intersection. He also emphasizes on the importance of keeping track of orientation and distance covered. These fundamentals are in sync with the framework of our application. He used our app on the second route. His ratings are as follows:

1. Time-distance understanding: Rating: 5/5

Comments: “I like the where-am-I feature. I would make my student use it multiple times till he got the hang of distance being covered. They have a tendency to stray off the route because it’s hard to walk straight lines. The orientation feature should be helpful.”

2. Accuracy of information: Rating: 5/5

Comments: None.

3. Usefulness of information: Rating: 4/5

Comments: “You have missed on a few useful clues. But I don’t blame you. Its hard to know if you are not an instructor. But the information you have, is helpful. ”

4. Availability of rectification information: Rating: 4/5

Comments: “The talking compass and distance measure is helpful when you get lost.”

5. Quality of training: Rating: 4/5

Comments: “You made a good attempt at using time distance, quizzes and activities. These are an important part of training. You should allow blind users to tag information too. A sighted volunteer may miss some important clues. But what you have is very helpful. In fact, if something like this was available in more locations, we could use it as a aide for training.”

4.3.2 User Experiments

Three of the six users had the assistance of a guide dog while all of them used a cane. The routes were chosen in a remote part of the campus and none of the participants claimed prior knowledge. We avoided doing the experiment in the night after some of our participants indicated their preference in learning a new route using the position of the sun during the daytime for guidance. We recorded the time taken, number of times assistance was used, techniques used, navigation apps of choice, etc. We asked the participants to use the iPhones provided in the accessibility mode. They could take breaks whenever they wanted and ask for assistance at anytime. We followed them at a distance to spot them for their safety. The participants were asked to rate the exercises on a 5-point Likert scale.

4.3.5 Problem Solving

The participants were asked to stop mid-training and perform a small subset of tasks. The tasks include going to the last mentioned landmark, pointing the phone at locations, figuring out cardinal directions and completing the assigned route. The users had to perform two tasks and two quizzes. Out of the six participants, only one of the participants failed to perform one of the tasks. Two of the users struggled to perform the tasks at the first attempt but with the assistance of the compass and WAI on the app, they could finish the task after a few attempts. This is an indication of the usefulness of the training application without human assistance. And the quizzes recorded an accuracy of 83%. The users indicated the quizzes contributed to their confidence on the unknown route.

4.3.6 Retention and Transfer

At the end of the experiment, the users were given a few tactile cutouts and asked to create the layout of the route. And then asked to point to the landmarks they remember in the approximate proximity on the tactile map they created. This tests their comprehension of relative locations of the landmarks. All the users successfully created the route. (While this was not to be a test of memory, each of the participants could name a dozen landmarks on an average.)

4.3.7 Experimenter's Remarks

The idea of stopping to do tasks and quizzes that are not a part of the route is something that can be implemented on campus for the following reasons. The pace on the campus is much slower than on a street. There is no traffic, which makes it safe to stop to evaluate the information. The automation of this training without a

human can also happen in a campus environment because of the absence of fear of being hit by vehicles. We observed that users who used guide dogs had a different take-home from our system. The dog is trained to look for an entrance or obstacles when the user stops on the street. We saw the guide dogs trying to steer the users to random places when the user stopped to take a quiz or needed to use the 'where am I' feature for additional information. Lastly, every user complained about too much information at the start of the experiment. But towards the end of the experiment, they were happy with the information provided. This suggests that it is acceptable to have rigorous training by giving rich information.

4.3.8 Scalability and Limitations

The success of this app hinges on the useful information that is available to the framework. As we found from the O&M trainers, there is a lot of useful information that a sighted volunteer who is not well versed with O&M techniques would miss. But the individuals with visual impairment would not miss these details. If we enlist a mix of volunteers and users to collect information over time, the app would become sustainable over larger spaces. Also, we tested our app on a campus that has grid like structure. Some older universities have buildings that are spread far apart with street traffic in routes. Adapting this app to a large space with buildings spread out would be extremely difficult. But the framework proposed translates well to small/medium size campuses. Initial data collection in this work was very tedious. We needed to poll coordinates at various landmarks from multiple volunteers to zero in on usable data. Creative crowdsourcing campaigns would be needed to scale for every campus. The users commented on wanting to add some

their own comments along a path, which might become useful notes for themselves later. This could be implemented by storing geo-tagged voice memos, which is not available in the current system.

4.4 Conclusion

We recognize a void in the requirements in learning a new space and the inadequacy in navigational applications in meeting these needs. We developed a novel approach to providing on-demand O&M training for visually impaired students on a college campus. Upon the development of an iPhone app implementing the idea, we evaluated its functionalities by real users in a campus. The O&M trainers who evaluated this app changed their mind from “not possible” to “good attempt”. The preliminary results are encouraging and demonstrate a large scope for further exploration along the same direction.

CHAPTER 5

TRANSLATING FLOOR PLANS INTO VERBAL DESCRIPTION

5.1 Problem Statement

Assistive technologies for the visually impaired have seen tremendous developments, producing innovative devices (e.g., various haptic-enabled devices and software products like widely-used screen-reading software). Nevertheless, there still remain many practical barriers for a blind individual who strives to lead an independent and active life. One such problem faced repeatedly in their everyday life is navigation, especially when visiting new places [10], [31]. White cane and guide dogs are the most commonly used aids for this purpose, last few decades have seen a lot of research dedicated to finding innovative technical assistances to help with navigation. Many of these technologies focused on GPS-based outdoor navigation or obstacle/landmark detection in indoor navigation. GPS-based navigation aids are among the most popular. These are indeed useful in outdoor environments but fail to give useful information indoors and in spaces where high precision matters. A significant amount of anxiety expressed by many visually impaired users relates to visiting a new place like a college campus and exploring new buildings like a library (twitter feed @blindperspective, [32]). Most of the readily available devices or software applications may guide a user to the vicinity of a building, but fail when finer details of navigating indoors are concerned.

Spatial abilities of individuals with visual impairment have been getting the attention of psychology researchers for a long time for both theoretical and practical reasons. Indoor navigation specifically attracted much research interest from

researchers in a quest to understand the nature of spatial representation. Recent years have also seen computer vision research being made to interpret map images for the visually impaired. Many mobile applications emerged, due to the increasing ubiquity of small mobile devices. There are also obstacle-detection-based devices built using sonar, optical imaging, and lasers. Bluetooth beacons and infrared signals have also been proposed in the last decade, which often require sensory devices to be strategically placed inside the building. Most of these existing methods do not help provide an overall sense of the location or help with planning a visit to a new location, nor do they help explore a new place. People still largely depend on seeking help from sighted people in the vicinity of the location to get help in finding places of interest.

Sighted people rely on visual aids when visiting public buildings like museums, libraries, college campuses, malls, hotels, etc. A floor map is usually available on the location or online. Providing an accessible map is an alternative that can be very useful in getting a sense of relative positions of locations and might reduce the need for totally relying on ‘asking for directions’. Consider a scenario where a student who is visually impaired wants to go to a local public library that he has never visited before. He could look up the details of key locations before he gets there or perhaps carry a braille print out or a voice memo that describes the locations of rooms inside the building. This might aide/encourage the student to independently explore the place since he has some prior knowledge about it. This kind of information presentation, aimed at overall cognitive understanding of a location, can be generated/delivered effectively in two different ways: as a verbal description, or as a tactile map (including an audio-tactile map). The tactile-map-based approach would

assume that the user has access to a tactile printer. Also, generating portable tactile maps of varying resolution and sizes containing textual information is very challenging to do automatically [33], [34], although simpler images like line graphs, bar charts, etc. can be relatively easy to process. Technical challenges aside, the low availability of tactile printers and the requirement of carrying many tactile maps may be additional practical difficulties. A more sustainable alternative is audio maps in terms of verbal descriptions of a place. The user can use this verbal description to get a sense of the space he would be entering and plan routes to access landmarks within the building. Hence an automatic indoor map/floor plan description generator could potentially greatly benefit the visually impaired community. A practically useful solution needs also factor into the design how a visually-impaired user demands verbal presentations of a place. This article reports an exploration towards developing an automated approach to generating useful verbal descriptions for indoor floormaps for the visually impaired.

The focus of this work is to develop a solution with existing technologies without relying on future product placement or human intervention. Navigation help can be provided in multiple ways. Useful information can be delivered in the form of orientation, details about the place, current location pin pointing, understanding of the nearby surroundings, help with selection of next point of interest, determination of route or overall understanding of the location of interest. *We present a prototype that intends to provide an overall understanding of the location of interest.* The first hurdle in building such a system to provide a description of a map is the availability of a usable floor plan of good resolution. A majority of public buildings have their floor map published on their Webpages. In spite of availability of a floor map, it

remains challenging if not impossible for a visually impaired user to go through all the Webpages, find appropriate links and assert that the image or PDF file is indeed a floor plan. Secondly, given an image of the floor map, the components of the map need to be determined automatically. Often, there is no standard way of representing the architectural details and using legend symbols. An automatic method to detect these details is a complex task by itself. Thirdly, assuming all the components of the map can be segmented, finding the right way to describe these components is another challenging task. To the best of our knowledge, there is no fully automatic solution for these tasks in question. This work is motivated by the fact that the blind users in our pilot studies raised the issue of unavailability of aids for indoor navigation. Yet another motivating factor is the observation that the majority of questions/problems raised on social networking sites of some of the prominent blind organizations are about directions once inside the building. This paper presents our attempts on developing a complete system for achieving the goal of automatically generating verbal description for supporting navigation inside a large indoor place. A preliminary version of the system was demonstrated on the ASSETS conference [35].

The proposed approach works around all the problems discussed above using a modular flow. We use keyword-based search and web-crawling to determine the availability of floor maps for a given place. If a floor map indeed exists, our system processes the downloaded image or PDF file to extract useful landmarks in the map. We then propose a simple yet efficient technique for generating a proper verbal description, taking into consideration of the complexity of the underlying map. As a focused case study, this work uses library floor maps as an exclusive example. Most

of the libraries make an effort to publish their floor maps online which makes it easier for us to focus on the development of the verbal description schema as opposed to finding the maps. The contribution of this work is three fold. First, a complete system was developed, which utilizes a simple interface and minimizes interactions required from a user. Secondly, automated map processing techniques were developed and integrated for factorizing the map into the necessary components needed for description generation. Thirdly, a simple yet efficient way to describe a floor map automatically is proposed, based on principles derived from pilot studies with visually impaired participants. The proposed system was tested on individuals with visual impairment by asking these users to recreate the map after listening to the description and comparing the rendering with the original map.

5.2 Relevant Work

The efforts put forth to alleviate navigation-related problems faced by individuals with visual impairment range from government policies, structural suggestions for buildings and sidewalks, to high-tech assistive technologies. Many street intersections have tactile markings and beeping traffic lights to aid with street crossing. The American Disability Act (ADA) standards [www.ada.gov] have a published standard for buildings to ensure ease of navigation for disabled individuals. Buildings constructed before the act often do not meet these standards. And most of the newer buildings meet the bare minimal requirement. This means entering an unfamiliar building remains challenging and sometimes even dangerous for a visually-impaired person. The law requires braille signs indicating room numbers to be provided at the door of each room. [36], [37] suggests physical

structural changes to buildings to aid navigation for blind individuals. All these efforts are a step towards helping with navigation, but still leave much room for further technological aids [38]. We provide a brief overview of some existing navigational technologies below.

Hardware-based technologies are designed to take an environmental input and provide haptic and/or audio output to the users. Most of these devices are designed to complement and not replace the cane or guide dog. A comprehensive history, evolution and reviews of some of these devices can be found in the work of [28]. We present a short summary in this section. They can be broadly categorized into sonar, optical, infrared and GPS-based devices.

Sonar-based: Sonar-based devices are a viable option due to their small size and hence the possibility of attaching them to the walking cane or personal mobile devices. The sonic image of the surroundings can be interpreted to find obstacles or to provide an auditory/haptic ‘map’ of the surroundings. Some of the available options in the market are GuideCane [www.ambutech.com], UltraCane [www.ultracane.com], Sonic PathFinder [www.sonicpathfinder.org], Sonic glasses, etc. One drawback of this technology is the need for the hardware, installation and maintenance. Secondly, there is a training stage to help the user interpret the output. The alerts given by the device and the intensity of the alert need to be constantly monitored to assess the obstacle ahead. This can be quite distracting. Also such devices are in general not very effective for crowded spaces.

Optical-based: Camera and laser based devices are the most common optical-based devices in the market. Numerous versions of laser canes have been devised and released in the market. More recently, computer vision community started

developing techniques to aid individuals with visual impairment to interpret surroundings using images. Methods [39] to create a 3D model from a collection of 2D images collected from the building in question have been proposed. This 3D model combined with localization techniques can be used to provide verbal directions. Common output is audio or haptic feedback. Older devices conveyed the information sequentially. For example, an audio signal was emitted pertaining to the information from the right to left along the vertical axis, in the photograph. A user needs to learn the image-to-sound translation rules. Interpretation of these rules gets more confusing as the complexity of the image increases. Newer devices use touch-based devices to provide better interfaces. Disadvantages with this type of techniques are the need for accuracy in image processing, typically long processing time and the fact that the lens has a line of sight.

Infrared-based: One of the most popular and notable devices using this technology is Talking Signs [www.talkingsigns.com], where infrared transmitters and IR receiver are used, with the latter attached to a user's cane. The system scans for signals in the proximity and emits an audio signal of the encoded message. Radio-frequency-based signage has become more popular due to low cost and ease of setup. One of the reasons a system like this is not more mainstream is the need to place these transmitters in the building and coding them.

GPS-based: To date, GPS-based systems are the easiest to use and hence very popular. Sendero GPS [www.senderogps.com], Trekker GPS [www.trekkerbreeze.com], Ariadne GPS [www.ariadne.com], etc. are some of the prominent players in the market today. These devices have done a good job translating the maps in an outdoor setting. But owing to accuracy issues, GPS

devices do not translate well into indoor-based navigation. Also, the requirement from the user side to carry additional equipment [40] like a PDA, wearable glasses, etc. has been a hindrance. This limitation can be overcome to some extent since most of the smart phones have this feature available in the form of a built-in map interface or as a downloadable application. Companies like Apple Inc. have taken a huge stride to make smart phones accessible, which has fueled the growth of newer technologies being based on the smart phone platform. However, aiding independent navigation for visually impaired individuals require information that is beyond the scope of what is readily available to these devices.

Software-based systems have become more prominent recently after the advent of touch screens, accessible mobile devices and innovative interface schemes. The prominent hardware GPS makers have applications that can be run on mobile platforms. We have seen maps become accessible in audio-tactile form [34], [41] and audio form [www.ariadnegps.com]. We have seen the advent of crowd sourced mobile applications like VizWiz [www.vizwiz.com] and 'Be My Eyes' [www.bemyeyes.org] enlist help of sighted individuals who answer queries. We have also seen map providers like Google and Bing provide accessible walking directions. Now the effort has extended to provide assistance in airports and shopping malls [www.bing.com/maps]. All the above-mentioned techniques have tackled specific areas of navigation on small datasets. We aim at exploring general methodologies and implementing a system for indoor floor maps.

When it comes to independent navigation, the most tested, accepted and commonly used methodologies for navigation for visually impaired individual can be found in the manuals of the Orientation and Mobility (O&M) training

[<http://www.caoms.org/>]. O&M training is a state-mandated training in America, which teaches safe navigation for everyday chores. O&M training provides the users with a skill set to equip them for safe and effective travel. O&M training has a strong emphasis on identifying and utilizing these cues to interpret the surroundings. Research has also been done of placing cues [42] in necessary locations and utilizing the signals to give on the fly information for assisting navigation and delivery of other information. iBeacons [<http://www.clickandgomaps.com/>] proposes the use of landmarks strategically placed in certain locations for navigation. This method, though effective, requires the system to be installed in the location and the user be trained to find the tactile markings. A step in the direction of building an aid to help the user to gain overall understanding of the space is work done on tactile map rendition with interactive feedback [43]. The limitation of such work is the need for the creation of a physical tactile map, which needs to be used in conjunction with a configured touch screen to access the corresponding text. In our approach, to build a system that works generically on any sort of map, we have tailored our approach to provide additional information to supplement the skills taught during O&M training instead of finding a stand-alone device-based solution.

We need to understand how spatial learning occurs in the human mind to create this solution. There are two ways to do this: through introspection – trying to record our own thoughts as they go by or through psychological experiments. Studies show that humans have two distinct ways of keeping track of orientation and position during travel [44]: landmark-based navigation and path integration. Landmark-based navigation uses visual, auditory, tactual, olfactory cues to form a *cognitive*

map for the navigator. Traditionally, the term cognitive mapping is associated with the mental image formed by humans when they traverse a path or a location multiple times. The map gets richer over time as the user accumulates more information each time he travels it. Non-visual senses like audio, smell, and object placement help enhance this representation. Multiple experiments have been performed to better understand the factors that contribute to better cognitive mapping.

Cognitive mapping is defined as a process composed of a series of psychological transformations by which an individual acquires, stores, recalls and decodes information about the relative locations and attributes of the phenomena in his everyday spatial environment [45]. Cognitive mapping research provides the baseline information necessary for an individual with visual impairment to explore a new space. The deficiency theory [46] states that lack of visual experience is an absolute obstacle for the development of appropriate spatial understanding. This theory has been proven inefficient by later works making way for the inefficiency theory. Inefficiency theory [47] states that individuals with visual impairment can understand and perform tasks based on spatial concepts but their comprehension is inferior to vision-based comprehension. The difference theory [48] argues that although nonvisual senses are at a disadvantage in providing spatial information, individuals with visual impairment have the same potential as their sighted counterparts to develop fully developed cognitive maps.

What is understood is that without vision, environmental interaction is hampered and that independent mobility requires greater skill [49]. The next aspect is to understand how a cognitive map translates to real time navigation. [50] used a

haptic environment to help the user form a cognitive map. And observed that the users were able to navigate new spaces with some success. They put forth the idea that alternate perceptual data may substantially support the cognitive processing of spatial information. [51] argued that the disadvantages faced by individuals with visual impairment is not due to the lack of visual experiences but the consequence of lack of proper coding strategies when it comes to receiving the information. These theories are crucial for our methodology. We do not claim to develop a tool for navigation but an aide to help the user to better understand an unfamiliar location.

Looking at the bigger picture, spatial learning before visiting a location is a technique that has been proven [52] to be a useful aid. We look into methods that aid the user to obtain an overall understanding without the use of tactile maps and by only using verbal description. This technique has been shown to be one of the preferred ways of communicating routes and landmark information [53] [54]. It has also been demonstrated that mental representation in visually impaired individuals is based on spatial relations and conditions described by words [55]. This ability to form a mental image in the memory is called “cognitive mapping” [56]. Spatial learning [57] and automatic route generation for indoor spaces [58] is a relatively mature research field that has not translated into any working system, partly due to the lack of digitized floor maps being readily available and lack of standardized dataset to work with. Work has been done on methods to generate verbal route descriptions [44], [59], [60] in general. Recent years have seen work [61], [62], [63], [64] on the same but tailored to effectively presenting this information to users with visual impairment. Though these are shown to be effective methods, many of them assume the availability of the map in a specialized format or need user intervention

in creating the maps, or even require assistance of a sighted volunteer in real-time. There is an existing disconnect between methods that make sense in theory but are challenging to use in practice for the end user. We propose an approach that does not make these assumptions, starts from a basic keyword search, works all the intermediate steps without manual intervention and produces a verbal description of the place as the output.

5.3 Proposed System

The prototype presented in this research focuses on making indoor floor plans accessible under practical constraints with today's readily available information. We make use of existing information for most public libraries and try to generate natural language text that is customized to visually impaired individuals. We present an overall system design in building a prototype system that works around all the problems discussed previously. Figure 24 shows the architecture of the proposed system. Each of these components will be elaborated subsequently. Although the current system may not be able to handle every kind of indoor maps that may exist, the general principles of generating content and natural language catered to individuals with visual impairment hold.

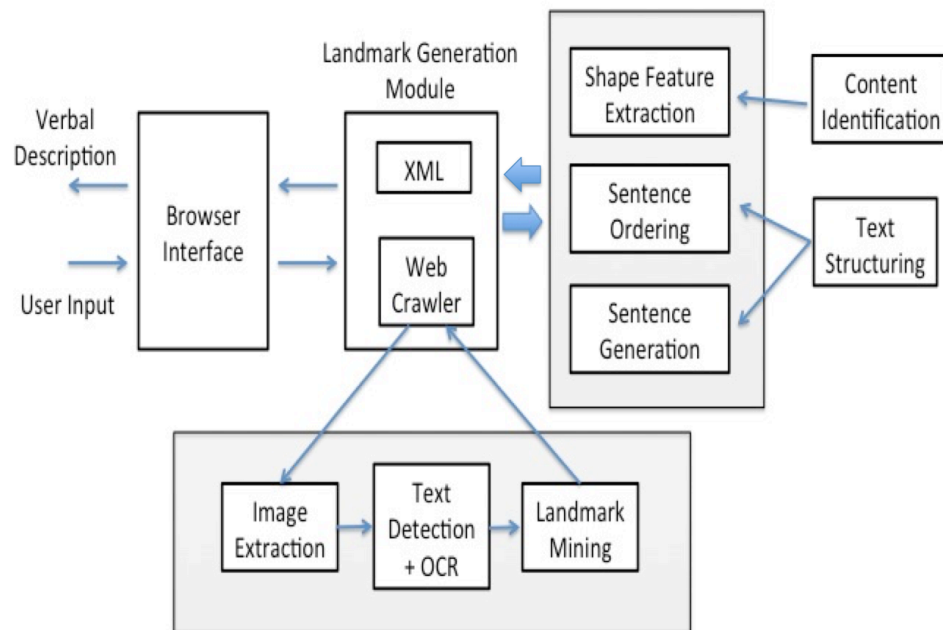


Figure 24 : The overall architecture of proposed system

The proposed approach to generating automatic indoor map description can be broken down into multiple interconnected components. The first problem faced is obtaining the appropriate floor plan given the name of the building. The user interacts with the Browser Interface module to enter the name of the location. This information is passed on to the Landmark Generation Module. These files are stored and indexed in our system in XML format. Recent interest in making indoor floor plans has resulted in the architectural details being available as CAD diagrams. Google Maps [<https://www.google.com/maps>] and Bing Maps [<https://www.bing.com/maps/>] have been steadily adding indoor locations that can enable navigation. Micello Maps [<http://micello.com/>] provides an API to add information into their database of maps. Availability of such information can enable us to bypass the Landmark Generation Module. But the reality is that there is still a need to obtain such information directly from many maps lacking such annotations.

In absence of indexed information, we invoke our web crawler to find us the required data online. A sighted computer user requires a couple of clicks on the website to find a link and download a floor plan from the library/hotel website. It is significantly harder for a user with visual impairment to browse the website and utilize the images found. Automatic floor map search/detection and extraction of keywords is the first step towards obtaining the data required to generate a verbal description. A sample of a potential input to our system is presented in Figure 2. In the absence of an available image or the indexed data, we return a prompt to the user telling him about the unavailability of the indoor map. The next step in the pipeline deals with pre-processing the map to extract useful information from the obtained floor plan. Filtering information that is most likely to help the user is a task by itself. We invoke the Content Identification module to re-purpose the data for the next stage. Once the content that needs to be presented is identified, we invoke the Text Structuring Module to generate natural language that is tailored to visually impaired users. The verbal description hence generated is sent back to the interface module that presents the information to the user. A sample illustration of the workflow is presented in Figure 25.

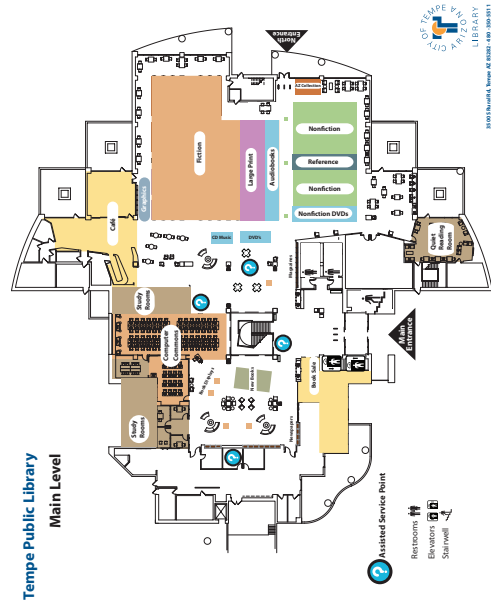
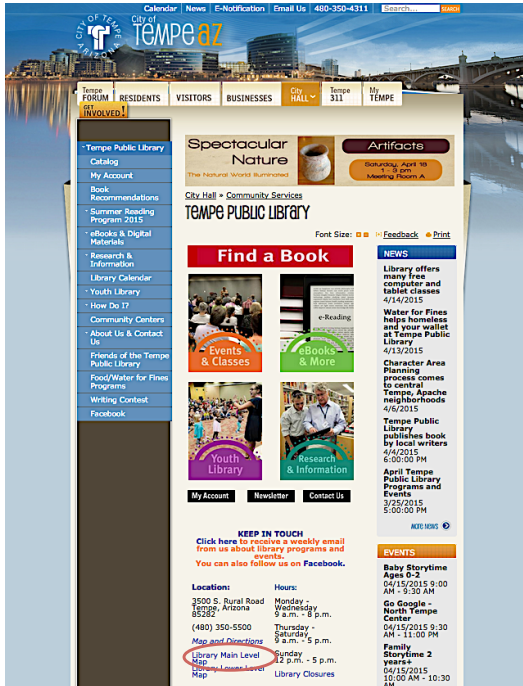


Figure 25 Shown to the left is the website of a public building. The red circle shows the link that redirects to the image of the floor map. Shown on the right : The floor map that the link leads to.

5.3.1 The Browser Interface

This module is in charge of interactions with the user and abstracting the back end modules. It receives the input from the user and provides the appropriate verbal description. This module is implemented using HTML+PHP, a server side script, for the following reasons. When an intense operation is performed on an image, it is taxing for a client side script to run it all. Secondly, most of these users run some sort of screen reader like Freedom Scientific's JAWS. We observed that running these two simultaneously as client side applications immensely slowed down the system. Chrome was the most popular browser of choice, as learnt from our case studies, and thus was used in this implementation. When a user opens our application, he is presented with a home screen where he can enter the name of the building he is interested in visiting. This is the entry point of the system. The user

has an option to enter the name of the building or the institution or enter the URL of the webpage of the location. We provide an auxiliary input method where the user or acquaintance of the user can upload the image of the floor map, as shown in Figure 27(a).

More often than not, we observed that institutions have more than one building with a floor plan. Sometimes, a building has multiple floors. We provide a secondary user input option where the user chooses the building/floor he is interested in, from the dropdown menu of available options, as shown in Figure 27 (b). Having the dropdown menu enhances the user control in generating the verbal description. This step is omitted if there is no choice to be made. The Browser Interface then accepts the input from the user and passes it along to the Landmark Generation Module. The Browser Interface also takes care of presenting the verbal description to the user. We have included the original image with alt tags for reference as shown in Figure 27 (c). The verbal description can be read out loud by the screen reader installed on the user's computer or a browser based screen reader. The description is presented in an HTML table. This layout enables the user to pause, repeat, read all, read line by line with the help of commonly used screen reader shortcuts.

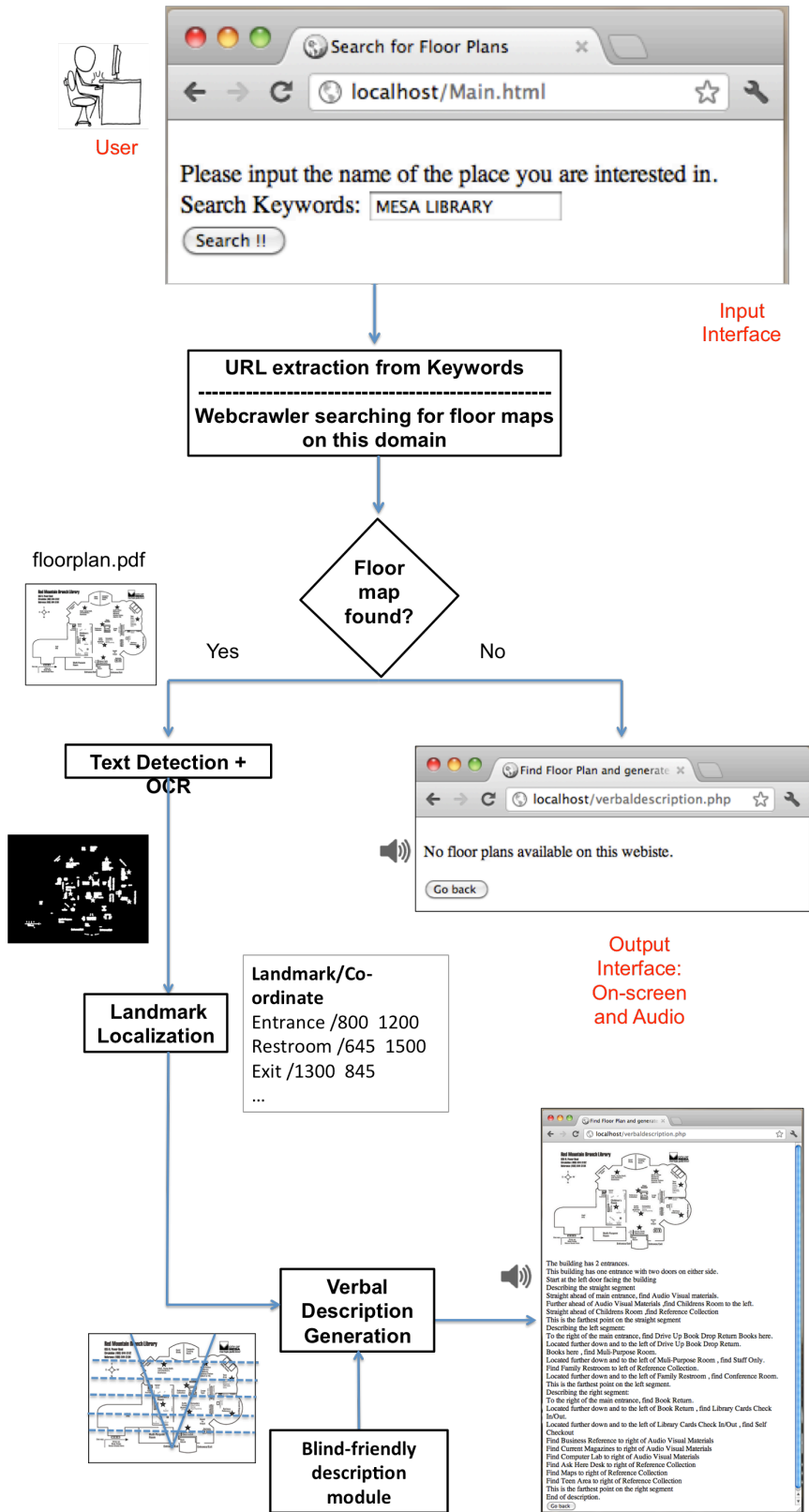
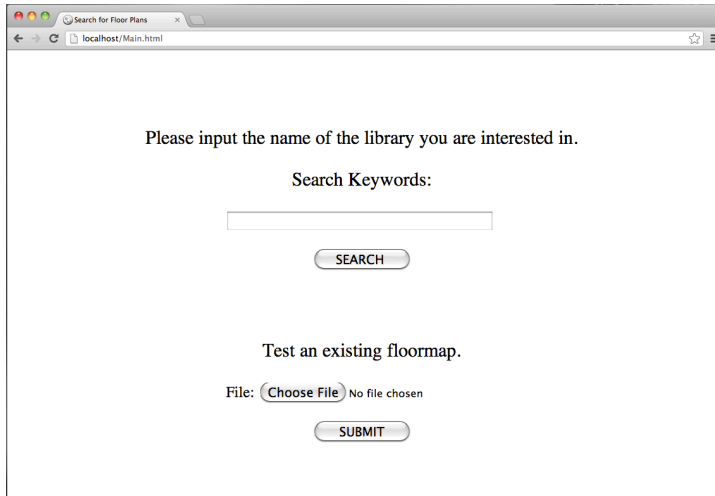
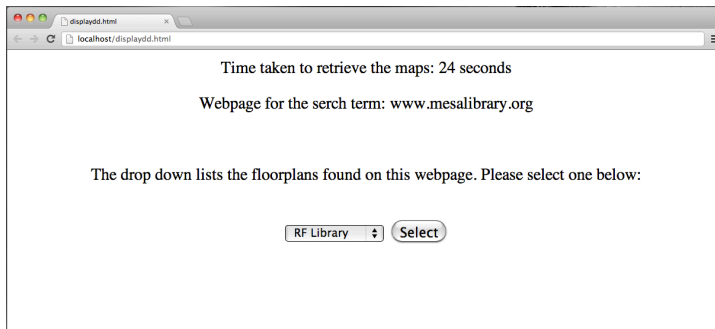


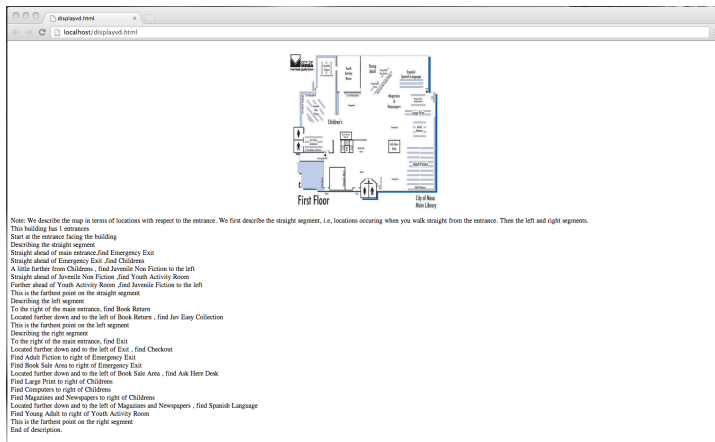
Figure 26 : An overview of the workflow of the system.



a)



b)



c)

Figure 27 : Input Interface. (a) The entry point for the user. (b) Confirmation from the user. (c) The final output.

5.3.2 The Landmark Generation Module

We define a landmark as any location of interest inside the building. We build this module to handle two cases. The first case being indexed landmarks from the source. The source could be a trusted mapping API like Google Maps or it could be straight from the building officials. With the increasing interest in building navigation paths inside large public spaces like airports, malls, etc., companies like Google and Microsoft are making an effort to acquire the finer details inside the buildings. This information can be accessed and stored for fast retrieval. The Landmark Generation Module takes in the name of the building as an input and generates an XML file containing the landmarks and their respective co-ordinates. There are multiple components in this module. We first look for availability of any such files corresponding to the location, in our indexed database. Efforts are in place to make files created using AutoCAD, geo-tagged data for prominent indoor buildings like airports and malls, available. Bing and Google Maps are utilizing this available data to integrate into their existing framework for location and directions. Although this is commendable undertaking, they currently have around 500 such indoor locations. We introduce this sub-step in our pipeline to cater to the future versions of this system where such information may be freely available. In spite of the efforts by corporations like Google, Microsoft, etc. to aggregate such information, most of the buildings are still not indexed. In such cases, we resort to our own search mechanism using a web crawler designed to work with the given constraints. This Web crawler looks for a website that publishes information about this building and looks for the corresponding floor plan on this verified website. If a floorplan is found, we proceed to extract landmark information from it. In the absence of such

information, this module lets the Browser Interface know that relevant information has not been found.

5.3.3 The Web Crawler Module

The Landmark Generation Module, in event of a corresponding indexed XML file not found, calls this module. It triggers a WebCrawler that uses the name of the building to find the host website that contains information about this building. We use Google's 'I am feeling Lucky' search API for this purpose. This feature has a near 100% accuracy on 50 libraries that we tested in finding the host website. We selected these 50 images from the first page of Google Images for libraries. Once we have a host URL, we invoke a PHP script that uses CURL libraries and DOM parsers to navigate every page originating from the host webpage to find graphic content. We look for keywords like 'floor map', 'floor plan', etc. in the link text in the HTML script as well as in the absolute URL for each link. We perform this search for all the links in the main page and continue this process for 3-5 layers depending on the search results on the fly and eliminating repetitions. We also remove the links that redirect to domains different from the domain of the main page. This step eliminates searching in hotels in different locations by the same group, other library buildings in the same institution, additional resources provided on the pages, etc. We 'follow' the URLs with relevant keywords till they lead to a page with a '.pdf', '.jpg', '.jpeg', '.png' and '.gif' extensions. We then determine if any of these files contain a floor map and download it to the server.

This naïve search technique is time-consuming. Hence we have introduced additional tricks to speed up the process. We noticed that libraries usually publish

floor plans in certain category of links like ‘About Us’/ ‘Building Information’ etc. We look for these terms on the homepage and check the links on these pages first. Most often than not, we have a hit within these links. In event of failure, we start searching the rest of the site. Also, avoiding searching pages that contain book databases, catalog information, employment information, calendar/event information, etc. has helped speed up the crawler. We have an upper time limit cut off and return a “not found” message to the user if the search exceeds this limit. We report a “No usable map found on this website” message when the floor plan published is an animated/interactive floor map, when the web master blocks crawlers from downloading content or when the website publishes a very low-resolution map.

5.3.4 The Visual Extraction Module

The Visual Extraction Module consists of three parts. First is the Image Extraction Module. The downloaded image is pre-processed and this module extracts the useful features. Second is the Text+OCR Module. We could not simply rely on any commonly available OCR detector for the following reason. When text is present inside and near graphics, additional treatment is required to get a good accuracy. We built a classifier of our own for this purpose. The third is the Landmark Mining module. This module generates a list of landmarks and their corresponding locations as an XML file for the subsequent modules.

5.3.5 Content Identification Module

This module is invoked by the Shape Feature Extraction Module to make sense of the data passed by the Landmark Extraction Module. As seen from our sample runs,

the XML file could contain many landmarks for a small building. This module identifies the salient and important features that will be utilized in the Verbal Description. We conducted a user survey and used the results of the survey to train this module. Features like shape information, terminology used, propositions used, etc. are recognized by the Content Identification Module. This module creates frameworks for sentence ordering and sentence generation. We use the incoming XML file containing the landmark information and the image of the floor map to “fill in” the required information.

5.3.6 Sentence Ordering Module

Sentence Ordering Module decides what should be the starting point in the image for the description, which landmark to pick next, and how to progressively select landmarks henceforth. The framework for this module is created using the results from user survey, by learning the preferred choice for such descriptions.

5.3.7 Sentence Generation Module

Given the landmarks and the adjacent details, we need a module to construct sentences for us. We could not readily use an existing NLP library because we wanted to tailor our output to the terminologies suggested by the O&M instructors and used by individuals with visual impairment. This module implement a framework to generate sentences based on the detected landmarks and the desirable terminologies and ways of describing a place most appropriate for the target population.

5.4 The Visual Extraction Module

We now elaborate some of the key computing steps in processing the downloaded visual map, before other further analysis tasks can take place. Figure 28 shows an overview of all the steps.

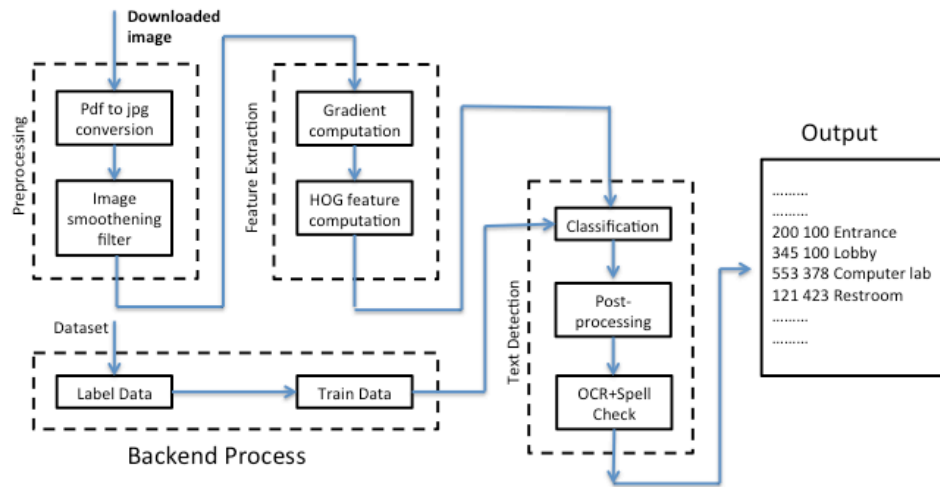


Figure 28 : Workflow of the Data Processing Module

Preprocessing on the downloaded images: As seen from the webpages of public buildings, image, portable document format (pdf) and interactive flash formats/animations are the preferred method to publish floor plans. Our current system does not handle animation files. If a map is in pdf format, it is converted into image using the pdf-to-jpg convertor from Boxoft (www.boxoft.com). In case of images corresponding to maps, color information is not very important for text detection. The image is converted into gray scale for further computation. Smoothing is done on the image with a 3 by 3-Gaussian filter to eliminate unwanted noise in gradient computations, while still preserving sharp edges.

Gradient based Feature Extraction: Gradient values along the horizontal and vertical directions at each pixel are obtained by gradient filters. From the obtained

values, the magnitude and orientation maps of the gradient are computed for the given image. Histogram of oriented gradient (HOG) features are then computed for text detection. We chose HOG features since the orientation distribution of text is considerably different from the non-text regions.

Text Classification and Location Registration: The first step towards building a classifier is to obtain training data. There are no existing datasets using floor maps, with ground truth available, for this purpose. We created a dataset of 50 floor maps from the first page of Google's image search on public libraries. We used a semi-automatic way of marking the ground truth by drawing demarcating rectangles surrounding each word in the image in the data set. The data corresponding to pixels falling inside these rectangles are labeled as +1 (indicating regions containing text) and those falling outside are labeled as -1 (indicating regions that are not text). The data are then used to train an SVM classifier for text detection.

Post processing for Improved Accuracy: The presence of lines and large connected components may result in noise. We detect lines and large connected components from the original image and create a mask with these regions. This mask is then applied on the SVM output to remove positively labeled pixels, which might belong to these regions. We similarly eliminate the false positives in the regions with zero gradient magnitude. We localize the text by detecting rectangles around each text region, and based on proximity, combine the words as needed. This results in improved text localization. More details of this process can be found in [65].

5.5 Content for Verbal Description

Human senses provide visual and non-visual features for understanding the environment. The visual features can be categorized into geometrical and non-geometrical features [62]. Geometric features pertain to the structure of the building. The non-geometric features could be lighting conditions, texture, objects for reference, etc. [66]. The problem on hand is: “What can be done to aide navigation in the absence of visual cues?” Studies show that visually-impaired users use non-geometric features to compensate for lack of visual cues. Giudice et al showed that using verbal descriptions of layout geometry help users navigate the space. They showed that pre-journey learning is useful for blind individuals when navigating a new environment. When working with cross-modality (graphical input and audio output), rather than assuming any mapping will work, a careful examination of how users present and interpret information is required.

Firstly, we need to decide what information (entrance, exit, orientation etc.) and how much of it (in terms of number of landmarks and geometric cues) to present. Secondly, we need to take into consideration a user’s experience and proficiency in understanding such information. Thirdly, the ordering and structuring of such information need to be determined. There is no single “how-to” guide or ideal way to achieve this. We adopt a case study based approach to derive the guidelines for doing the mapping from images to meta data to verbal descriptions. This section elaborates on such case studies, the findings and the conclusions.

5.5.1 Deficit with Existing Models

Generating useful verbal description for a given spatial layout is a rather challenging task. [67] uses corridor layouts generated with the help of an experimenter using software with a user interface for the user to manually mark out some of the necessary inputs. Most of the maps as seen in Figure 5 cannot be broken down in corridors since they have a more generic layout. Hence description in terms of intersections is not the most appropriate way to describe these maps. Often verbal descriptions involve relative positions given the current location, relative positions in terms of the distance, directions in terms of cardinal directions and directions in terms of absolute directions and distance. Without assuming the availability of scale information and the map orientation, much of such information is hard to automatically obtain directly only from a downloaded map image. In particular, directions in terms of absolute orientation need the user to face a certain direction and keep track of orientation all along the route. Kalia, et. al. proposed and tested two types of description for navigation. First method is called allocentric, where the description is absolute in terms of cardinal directions. The second is called egocentric method, which uses the user as the point of reference and describes every other location as a relative position from the user. We do not use the allocentric or egocentric descriptions proposed, because the problem we are trying to solve deals with describing the entire spatial layout, as opposed to taking a user input in real time and describing the layout around that location.

5.5.2 Learning from our Target Audience

In the absence of a standardized “how-to” guide when it comes to giving spatial information verbally, we resort to analyzing human written text messages and using the domain expertise of the orientation and mobility instructors, for this task.

We performed a small case study and tried to learn how blind users verbally describe a building that they are familiar with and also tried to learn how they usually receive layout information from sighted individuals. We asked a small group of four blind users (average age = 27, range = 21 to 52), to describe a building on campus that they usually need to go to use the accessible textbook materials and are very familiar with, to another blind individual who is new to the building. We chose this building for the following reasons: the building has four entrances, 20 landmarks that are of interest and has a rectangular structure (which is a common shape for buildings). The three students and one professional who participated worked in this building, and access it regularly. Two of the participants were born blind and the other two lost their vision during their teens. The users were given complete freedom to describe it as per their preferences in terminology and schema. They were given two days to complete the assignment. We were in turn asked questions like: “do we need to start with a particular entrance”, “should I describe every location in the building”, etc. We asked them to describe it the way they find most useful to a fellow visually impaired student. The second stage of this study involved critiquing the descriptions and rating them by usefulness. We made a list of all the descriptions and distributed them to all the participants. For the sake of reference for the participant, we included his/her version as well. To this list, we added a description obtained from a sighted person who works in the same building.

This was included to learn the kind of criticism it would face, in comparison to the rest of the descriptions. The participants were asked to comment on each description and rate it on a scale of 1-5 with respect to clarity, usefulness and accuracy. We used this feedback to arrive at paradigms to be used for generating verbal description. The case study was inspired by [68], where they had asked the residents of Venice to provide descriptions of routes in the city and used this data to analyze how humans model descriptions. Some of the sample responses are provided below :

Participant 1(highest rated, average=5/5): “The Disability Resources Center building is a rectangular building with the long sides going from north to south. When you enter from the north door, the door to your almost immediate left will be the entrance to the room where most of the braille transcribing is done. Across from this door on the hall's west side is the office of the supervisor of this department. Moving southward down the hall, one comes across an open study and get-together space on the left with a door leading to the outside with steep steps almost immediately after opening it. Across from this open area is a window with a small hall just south of it that represent the location of the testing center. Moving further south down the main north-south corridor one comes to a flight of stairs going up to the second floor on the left. A little further south on the right is the men's bathroom labeled in braille followed by the break/lunch room (at its west end there is a door leading out to an enclosed patio) followed by a drinking fountain followed by the door to the ladies' bathroom also labeled in braille. Meanwhile, the left-hand side of the main hall has doors leading to various private offices and a meeting room. Beyond the ladies bathroom on the right as you continue down the main hall is another hall that is

very important. It leads back to the computer lab (on its north side) and the offices of the various disability counselors (south side). Returning to the main hall and continuing south, one enters a relatively large room where the two secretaries sit. The main desk if you have questions is directly in front of you. The other secretary sits east of the main desk and beyond her is another door leading to the outside at the southeast corner of the building. Across from the second secretary's desk on the north side of this room are doors that lead to first the office of the director's private secretary and then, beyond that, the office of the director of the department. Finally, next to the main secretary's desk on the right is a closed door that leads to the Transportation room.”

Participant 2: “The building is a rectangle with the longer side along North-South. It has four entrances. The main entrance is the easiest to use for description, hence I want to use this. When you enter through the main entrance, you will find a waiting and tutoring hall with tables and chairs. If you walk further, you will find the receptionists desk. From here, you can go straight, left or right. If you go right, you will find a hallway and the Alternative Formatting Center to the right and the supervisors office to the left. If you keep going further, you will find the Northern entrance. If you go straight ahead from the receptionist’s desk, you will find the Testing center supervisors office to your right. And further ahead, you will find the testing center. If you take a left from the receptionist, you will enter another hallway. You will find a meeting room and some offices to your left. To your right you will find the men’s restroom, a common purpose room and the women’s restroom. Going further, you will find that the hall way widens. Here, to the right,

you will find a computer lab and offices of DACs. If you continue along the hallway, you will find another reception area, the transportation center and the South-East entrance.”

Participant 3: “If you enter the DRC from the southeast entrance you will enter the reception area. This is set up like a waiting room might be. There will be a receptionist desk straight ahead of you and another desk to the right. To the left, there are several sofa chairs and tables. If you walk between the desks and head north, you will be walking down the main hall of the DRC. There are offices on either side as well as the transportation center behind you at the south side. If you walk down the main hall, you will notice an opening on the left that leads to the lab and other offices. If you keep going, you will approach the bathrooms and drinking fountain on the left. The door directly after the girl's bathroom is the multi-purpose room. If you keep going you will notice another opening on the left. This is the testing center. On the right is a lobby with tables set up for tutoring. There is a walkway between the tables leading to a door on the northeast side of the building. If you continue down the main hall you will pass the alternative format center on right and the north entrance to the building.”

5.5.3 Case Study Findings

The more blind users we included in the experiment, the more diverse answers we got. A sample description given by one of the female participants is given below:

“If you enter the DRC from the southeast entrance you will enter the reception area. This is set up like a waiting room might be. There will be a receptionist desk

straight ahead of you and another desk to the right. To the left, there are several sofa chairs and tables. If you walk between the desks and head north, you will be walking down the main hall of the DRC. There are offices on either side as well as the transportation center behind you at the south side. If you walk down the main hall, you will notice an opening on the left that leads to the lab and other offices. If you keep going, you will approach the bathrooms and drinking fountain on the left. The door directly after the girl's bathroom is the multi-purpose room. If you keep going you will notice another opening on the left. This is the testing center. On the right is a lobby with tables set up for tutoring. There is a walkway between the tables leading to a door on the northeast side of the building. If you continue down the main hall you will pass the alternative format center on right and the north entrance to the building.”

We summarize our observations from the responses from all the participants below:

1. Navigation bias: From our acquaintance with this set of participants, we could tell that they were describing their everyday path taken, when they access the building facilities. In spite of being asked for an overall description of the building, the users tend towards giving descriptions geared towards navigation. Their preference in entrance, usage of specific rooms and interaction with the building staff is evident from their descriptions.
2. Gender based observation: One of the female participants only used the relative positions in terms of “left/right/straight” to describe the landmarks inside the building while the two the male participants used a lot more cardinal directions instead.

3. Visual Disability: The low vision students did use more landmarks when compared to blind students. However, we observed that there were not a lot of obvious fundamental differences between low-vision and blind students in their descriptions.

4. Common vocabulary: Commonly used words for descriptions are “to the left/right of”, “straight ahead”, “across the room to the left/right”, “down the xyx, find abc”, “going further”, “immediate left”, “down the hallway”, etc.

5. Detail: The participants wrote the description using all the landmarks they could remember. They did not leave out any landmark that they deemed unimportant or unnecessary. They used references to non-structural landmarks like tables, chairs, desk, receptionist, etc. to make it more explicit.

There is no one universal way to describe the building that can be arrived at from looking at the descriptions. But we did notice certain “trends” when it came to how they framed their descriptions. We conducted a second round of experiments where we asked the participants to rate and comment on each other’s descriptions. Our observations can be summarized below:

1. One of the key comments from the participants was about the use of cardinal directions, where some users indicated their discomfort in being able to figure out directions inside the building unless they concentrated on every step during their journey. In spite of this comment, every participant started his or her description at one of the entrances of the building, described by its cardinal position (e.g. North entrance).

2. A detail that got good feedback from the participants is the description of the overall shape of the building. Another detail that the participants pointed out was

the usage of words like “immediate”, “few steps” and “right after”, that convey distance information.

3. An interesting observation is that the participants critiqued their own description and pointed out possible improvements after reading the rest of the entries. They acknowledged the difficulty in coming up with a good description of the building and admitted that their descriptions may be “confusing to a first time visitor”.

4. One of the most criticized details pertains to a certain description using “start at the main entrance” without describing the location of the entrance with respect to cardinal directions or the position with respect to the building.

5. The description given by the sighted volunteer got the lowest rating (average- 1.5) and most criticism of all the included responses. The participants complained about lack of details, incorrect directions and described it as confusing. One of the participant blatantly mentioned that this person is confused or lacks proper direction sense. This interestingly underscores our approach of learning from the visually impaired users to develop a scheme for our automated solution.

6. The participants were surprised when they read about the existence of landmarks that they never knew prior to this experiment, in spite of being familiar with the building.

In spite of the diversity in descriptions, the ratings given to the descriptions and the participant feedback provide a conclusive indication to the direction to follow in order to generate the most effective paradigms needed to describe a building. We derive our design principles after analyzing such user-based inputs.

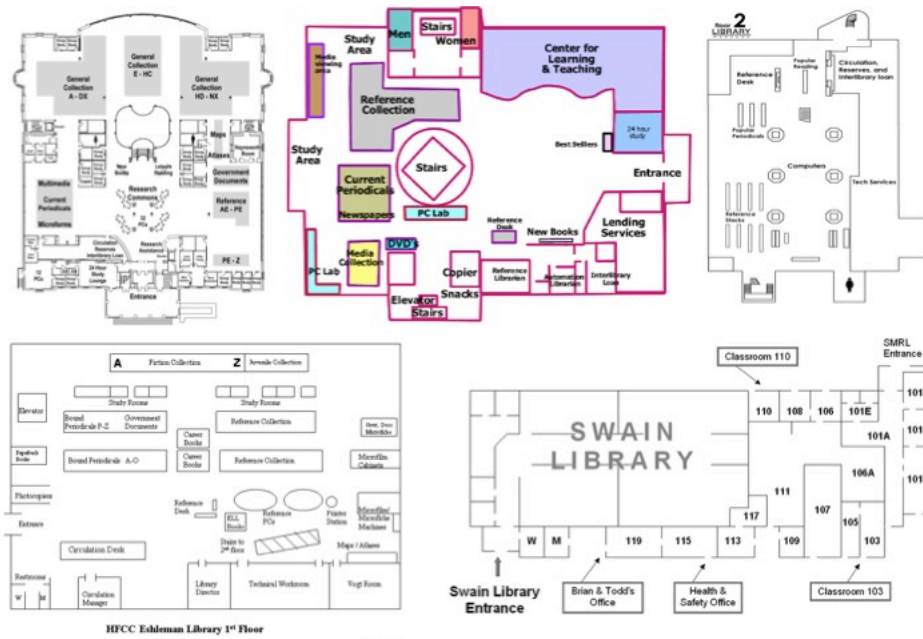


Figure 29 : A few sample floor plans found on the web using Google Image Search

As seen in Figure 29, the images found on the Internet come in various kinds of representations. We seldom see a standard way to representation that is followed by the publishers. In almost all of the maps we have seen from Google’s Image search, none of the maps have a compass indicating the cardinal directions or contain information in terms of hallways. Some of them contain information about the smaller landmarks like the tables, chairs, doors, etc., but do not follow a standard format. A disadvantage from the dataset is the fact that in spite of the presence of above-mentioned important information, we cannot use it to train an automated system because of the diversity of the maps. For example, it can be seen that a circle is used to indicate a computer table in one of the maps in Figure 6, and denotes a stairway, a lobby and a circular building structure behind the stairs in other images. Generalizing an algorithm to detect these details may give wrong information that

could do more harm than good. Owing to these limitations, we concentrate on the textual cues contained in the map and base our description on it.

5.5.4 Paradigms for overall description

After a detailed analysis the descriptions and interaction with the participants, we tried to mimic the observations and the feedback, in designing the verbal description generator. The summary of the observations is given below in terms of some general guidelines for producing useful verbal descriptions.

1. Shape: Start the description with the overall shape of the building. Describe the available entrances to the building with reference to the shape and the cardinal directions (if this information is available).
2. Start point: Use the entrance of the building as the reference point to start the description of the landmarks inside the building. We should detect the text for entrance or in their absence, symbols like doors/double doors that could potentially be the entrance to the building. We can then start at this location and base rest of the descriptions form this point. For floors that are not the base/main floor, we can detect the elevators and start at an elevator.
3. Entrance: From the study, there is no one entrance that can be deemed the best reference when more than one entrance is available to access the building. The safe bet is to start with the entrance that is labeled the “main entrance” or the one leading into the center of the building.

3. Flow: Most of the users describe the locations as they enter the main door and describe what is to be found as you walk along. We start our description with all the locations found when navigating straight from the main door. They then described what is found to the left and to the right at intersections. This scheme has many outliers when dealing with non-rectangular layouts, which we explore in the subsequent sections.

4. Position: The users frequently used “to the left” and “to the right” to describe locations with respect to a previously mentioned/described location. We should describe rooms/landmarks in terms of their relative position with respect to a location already described in a previous step.

5. Distance: From the descriptions, the only indicators to the distance was the usage of terms like “to the immediate left/right” and “further ahead”, indicating proximity and large distance, respectively. We use a similar scheme to convey the proximity of locations.

5. Popular references: All the users included landmarks like lobby, reception, restrooms, stairs, elevators, lobby, etc. Accordingly, we created a small dictionary of building-related terminologies that is used to define important landmarks.

6. Segmentation of space: The building used for the case study is a rectangular one with long hallways running across the length and breadth of the building. The users

based their descriptions on the landmarks encountered walking along these hallways. They branched away from the main hall way and described location encountered along the branches of the hallway, thereby, dividing the space into segments. This was not done explicitly but can be observed consistently in the descriptions. We use a similar scheme to segment the space.

7. Categorization of the essentials: A few participants included finer details like the position of pillars, number of steps, position of the tables in the hallway, etc. in their descriptions. Since this level of details is not available without real-time indoor imagery of the space, we chose to ignore these details. On the other hand, the details of positioning of hallways and stairs, location of restrooms and water fountains, etc. are essential. We give utmost importance to delivering such information and pay attention to the format of this information to best deliver the message.

The paradigms arrived at are very coarse in outlook. This was necessary to generalize these rules to different layouts that could exist. We asked the participants to describe a rectangular layout, which has four entrances. Most of the users wrote that they picked the entrance that leads into the longer side of the building because it was easier to describe. Looking at the descriptions, it is evident that the participants, in spite of being familiar with the building, have not explored it fully. This also points out to the need for a verbal description system. Perhaps having them explore the map would give them more confidence to find new locations and trace their way back. The landmarks that everybody used in common are the restrooms and lobbies. We try to emphasize these landmarks in our descriptions and

use them as major reference points. The proposed paradigms address the different aspects involved in describing a physical space. We arrived upon a rule for each of these aspects that work for the common layouts and can be extended to complicated layouts with reasonable accuracy in depiction.

5.5.5 Lexical Choices for Describing the Concepts

One of the major inferences from the descriptions from our participants is to learn the terminology in use. This information coupled with feedback from the Orientation and Mobility Instructor, helped us narrow down on the terminology to use, as given below through examples.

1. Shape: The participants used the words “longer/short hallway” to express some sort of shape information. We include shape information as the first step in the verbal description to improve upon the overall understanding of the layout of the building.

- a. Shape (square, rectangle)
The building is square in shape.
- b. Side (length, breadth)
The longer side is along the breadth. The shorter side is along the length.

2. Entrance: In spite of not being asked to give directions to a certain location inside the building, every participant started their description with the entrance.

- a. Number of entrances (building)
The building has two entrances.
- b. Located (along, length, breadth)
Both the entrances are located along the length, one each to the right and left of center of the length.
The entrance is located on the breadth, towards the center.

3. Position: As seen from the descriptions, the users had a sequential way to add on more landmarks into the description. The most salient feature of positioning is relativity.

- a. Adjacency
Adjacent to location A, find location B
- b. Egocentricity (left, right, straight, behind)
Find location B straight ahead

4. Distance: Pack the description with consistent pointers to let the users know distance that is not quantified but still describes proximity or afar.

- a. Proximity emphasis
Find location B to the immediate left.
Location B is also located straight ahead of location A2.
- b. Convey if not near
To the farther left of location A1, find Location B.
Location B can be found moving further ahead from location A.

5.5.6 Rules for Content Determination

Using the classification of the essential information and necessary aspects of generating the description, we define six high level rules to serve as the framework for the description. Some of these rules are contingent upon the other rules. Some of them are stand alone rules. We included the rules below. For illustration purposes, we picked out a statement each from our real-time descriptions generated by the system so far.

- 1. If (enough descriptors have been successfully detected), then include the shape information for the building.
The building is rectangular in shape. The longer side is along the breadth.
The shorter side is along the length.
- 2. If (cardinal directions available), then add (facing direction) to the subsequent statements.
- 3. If (more than one entrance found), then add (overall entrance information).

The building has two entrances.

- a. If (more than one entrance found), then add (relational positions).
Both the entrance are located along the length, one each to the right and left of center of the length.
- b. If (cardinal directions available), then add (relational directions).
The entrance is located on the North East corner of the building.
4. If more than one entrances found, look for the terms (main, primary) in the text. If none found, then start at an entrance that is located along the shorter length of the building.
Start at the left entrance facing the building.
5. Start with the entrance, and pick a location near the entrance based on its relative positioning. Use relational articles to generate sentences.
To the right of the entrance, find Circulation Desk.
6. Inculcate distance information in terms of (a little further, straight ahead).
Straight ahead of Reference Desk, find general reference.

5.6 Verbal Description Generation

Generating an overall verbal description for a spatial layout is an area of research that has not gained enough attention. Customizing this description to facilitate spatial reconstruction for visually impaired users is a trickier solution to arrive at when dealing with diverse layouts. Using our derived paradigms from our case study, we propose a description generation method that is aimed at providing the user with an overall sense of the space under consideration.

5.6.1 An Overview

We propose a sequence of steps to be followed to generate the verbal description. Each of these steps will be described in detail in this section. The algorithmic overview is given below:

1. Step 1: Understanding the geometry of the building

- a. Describe the overall geometry and the location of entrances to the building with respect to the overall shape of the building.
 - b. Determine the number of entrances to the building. Pick the appropriate entrance to be the starting reference point.
2. Step 2: Segment the map into regions.
 - a. Create the appropriate number of segments based on position and the density of the landmarks.
 - b. Based on the dimensions of the map, create a grid that can serve as a reference for the proximity of landmarks.
3. Step 3: Create a structured flow for verbal description
 - a. For each landmark, find the appropriate pairing for reference using the segment, grid and proximity information. Call the function to generate verbal description given this information.
 - b. Determine the next point to describe. Repeat.
4. Generate verbal description
 - a. Create a set of words to use in a sentence to describe the relation between any two points from the case study.
 - b. Generate the sentence based on the pre-defined word set and the information about the segment, grid and proximity.

5.6.2 Overall Geometry Understanding

The first step of the verbal description process is to give the user information about the overall shape of the building. The most commonly seen building shape is a rectangle, square or a rectangle with an extended wing on one of the sides. From the data acquisition module, we have the locations of each landmark in the building. We present one of the layouts from a local library in Figure 30. The shape of the building contains an arbitrary number of crevices and extensions out of the rectangle. The aim of generating description of the overall shape of the building is not to convey the intricacies of the layout, but to convey a general idea of the space enclosing the landmarks. The example shown below has most of the landmarks inside a rectangle. The building has a small wing at the bottom right corner of the rectangle containing two landmarks. Generating this information was deemed very useful during our case study. We use the entrance as the reference point to generate

this information. We generate a rectangle touching the entrance and enclosing the landmarks in the direction of the centroid of the building structure. We check for the corner points of the rectangle for graphics to make sure it is enclosing the building or adjust the rectangle accordingly. We then check for outlier points to look for any wing extensions the building might have. Our verbal description consists of information regarding the main rectangle, the location of the wings and the position of the entrance with respect to the overall geometry that we previously described.

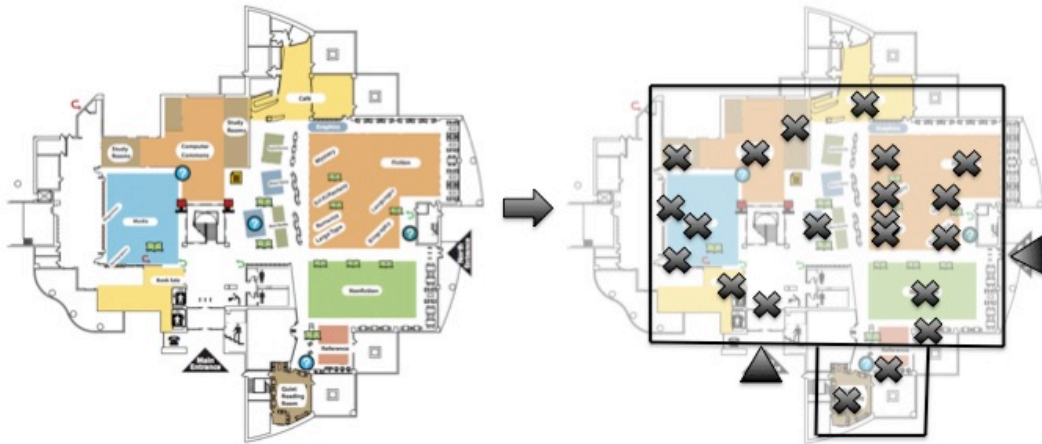


Figure 30 "This building is rectangular in shape which is longer across the length or in the horizontal direction. The main entrance is located at the bottom center of the rectangle."

The most natural way to describe a building as seen from our case study can be summarized as follows: describe the rooms or landmarks encountered when walking straight into the main entrance door. Then describe locations to the left and to the right either with respect to the objects already described previously or with reference to the entrance. As an attempt to simplify this task, we divide the map into three segments: 'straight', 'left' and 'right' segment, as shown in the Figure 31. The straight segment does not mean locations encountered when walking in an exact straight line. Studies [41] have shown that the sense of 'straightness'

decreases as you go farther from the original point of reference, for an individual with visual impairment. To account for this fact, we used a small angle on either side of the line tangent to the entrance to define the straight segment. The remainder segments are divided into the left and the right segments. Sometimes, the entrance is not the outermost point of the building; we add the outlier locations to the segments accordingly. We divide the entire floor into grids to quantize the location of each landmark to act as a point of reference when conveying the distance information. A sample illustration can be found in Figure 31. More often than not, segmentation is not a straightforward division into three parts and a grid. We have considered the geometry of the building, the aspect ratio of the length and breath of the building to define the most feasible scheme for description. The details for this procedure are elaborated in the next section.

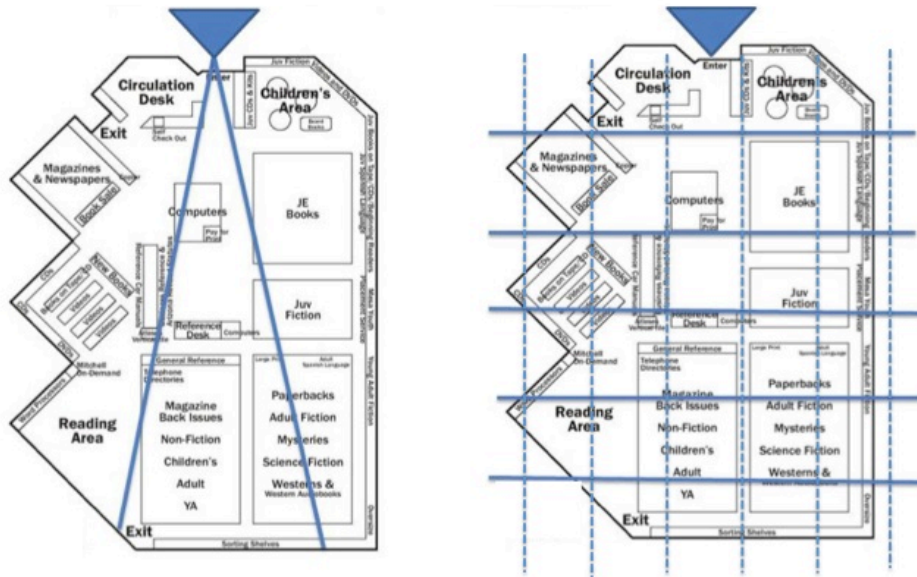


Figure 31 The image above illustrates the segments used to divide the map.

We follow the steps described below to sequentially scan through the points and determine the relative location to describe every point with. Since there is no

standard way to specify the entrance in the form of a symbol, we use textual cues to figure out a logical sequence in picking a new point to describe and the point to use for reference for relative location. We divide the map into segments and grids; form an intermediate spatial reference to help for a structure of reference points that need to be described with respect to one another. These steps can be summarized as follows:

1. **First step:** Detect an Entrance
 - a. Look for the words 'entrance', 'main entrance', 'main door', 'enter', 'exit' in the text detection results.
 - b. Entrance count = Count the number of entrances;
 - i. Output: "This building has" Entrance count " locations marked as entrances"
 - c. Look for 'main' to determine the entrance to use for the reference.
 - d. If no 'main entrance' found and more than one 'entrance' found, find the outermost entrance.
 - e. Determine the direction of the map with respect to the entrance
 - i. Use the angle between the centroid of the map region and the centroid of the bounding box for text containing the keywords for entrance.
 - ii. Output: "Start at" main entrance "facing the building".
2. **Second Step:** Pick the quadrants containing the map treating the entrance as the origin.
 - a. Three regions: Draw 90-ed tangents in the direction of the map. If all the text centroids fall in the region, 3 regions
 - b. Four regions: else check 3rd and 4th quadrants. Include the quadrant as needed.
 - c. Return the number of quadrants.
3. **Calculate theta:** Determine the extent of each segment
 - a. Square building
 - i. Entrance approximately located at the center: $\theta = 20$.
 - ii. Entrance closer to the edge of the building: $\theta = 15$.
 - b. Rectangular building, entrance located on the shorter side.
 - i. Entrance approximately located at the center: $\theta = 15$.
 - ii. Entrance closer to the edge of the building: $\theta = 10$.
 - c. Rectangular building, with entrance located on the longer side.
 - i. Entrance approximately located at the center: $\theta = 22$.
 - ii. Entrance closer to the edge of the building: $\theta = 15$.
4. **Third Step:** Pick the quadrants contained in the tangents.
 - a. Add outlier points as needed.
 - b. Divide into straight, left and right regions.
 - c. Use theta degrees leeway for 'straight' segment.

- d. Segment the rest into the left and right segments.
- 5. **Create a grid:** Divide the map into a large grid to capture the proximity of text regions from the adjacent segments
 - a. Calculate the length and breadth of the building.
 - b. Create a grid across the length and breadth of the building. We used a 100 square pixels per square rule to create the grid.
- 6. **For each segment:**
 - a. Scan by the lowest to highest row or column depending on the position of the entrance.
 - b. Pick the next point pair by scanning sequentially along the segment.
 - c. For each point, look for additional points in the proximity to describe.
 - d. For left/right segments: find a point of reference from the center segment. If none found, use a previously described point in the same segment for reference. Use the appropriate terminology based on segment, proximity and direction.
 - e. Call the sentence generator once the parameters: new landmark, point of reference, segment, proximity and direction, are calculated.

We divided the map into a spatial grid that's intuitive to describe each point wrt another. We first describe the points located in the central segment, followed by left and right segments. We repeat this procedure going further along the grid we created. The verbal description for the straight segment uses a previous reference, for example "Find Computer room ahead of the main entrance". We use the information about the grid to give spatial cues about distance. If two locations are contained in the same grid, in the straight segment, we mention the relative position in terms of left and right, for example, "reference desk located to the right of the computer lab". Locations located on a different grid have the word 'further' to emphasize the relative position, for example, "study room located further ahead from the computer lab". The left and right segments have a slightly difference referencing method. We use a point on the same grid in the straight segment that has previously been described as a reference to each point in the left/right segment. If no point exists in the straight segment, we use the previous point in the segment to generate the spatial relation, for example, "magazines located to the right of reading area".

We apply this simple procedure to all the points to generate a complete verbal description of the place. All the scenarios are illustrated in the Figure 32. For the sake of brevity, we do not repeat figures to illustrate subtle differences in description from the point being located to the right instead of left or vice-versa. We use one case to illustrate our sentence generation for this scenario.

When applying these rules to a large set of test images, it is not uncommon to find images that are a hit or miss on criterion we use as anchor points. For example, if we fail to detect an entrance, we exit the program and inform the user. We encountered maps of very low resolution where everything is noisy. In this case, we still go ahead and do our text detection. If the dictionary can not match more than half the words, we inform the user that this map is not usable. We have also seen maps of extremely large spaces with hundreds of landmarks. In this case, we matched the locations with the dictionary, and tried to eliminate some of the landmarks. In the future versions, we plan to build an interactive interface that lets the user weed out the excess.

We show some of the pairs of points and their respective description. The red marker indicates a landmark (location A) that has already been described. We use A1 and A2 to indicate using two previously used landmarks for reference. The black marker indicates a landmark (location B) that is to be described next. (a) "Adjacent to location A, find location B to the immediate left." (b) "Further ahead from location A, find location B straight ahead." (c) "Further ahead from location A, find location B to the right." (d) "To the left of location A, find location B." (e) "To the left of location A1, find location B. Location B is also located straight ahead of location A2." (f) "To the farther left of location A1, find location B. Location B is also location to the left of

location A2.” (g) “Location B can be found moving further ahead from location A, to the immediate right.” (h) “Location B can be found moving further ahead from location A, straight ahead.” (i) “Location B can be found moving further ahead from location A, to the left.”

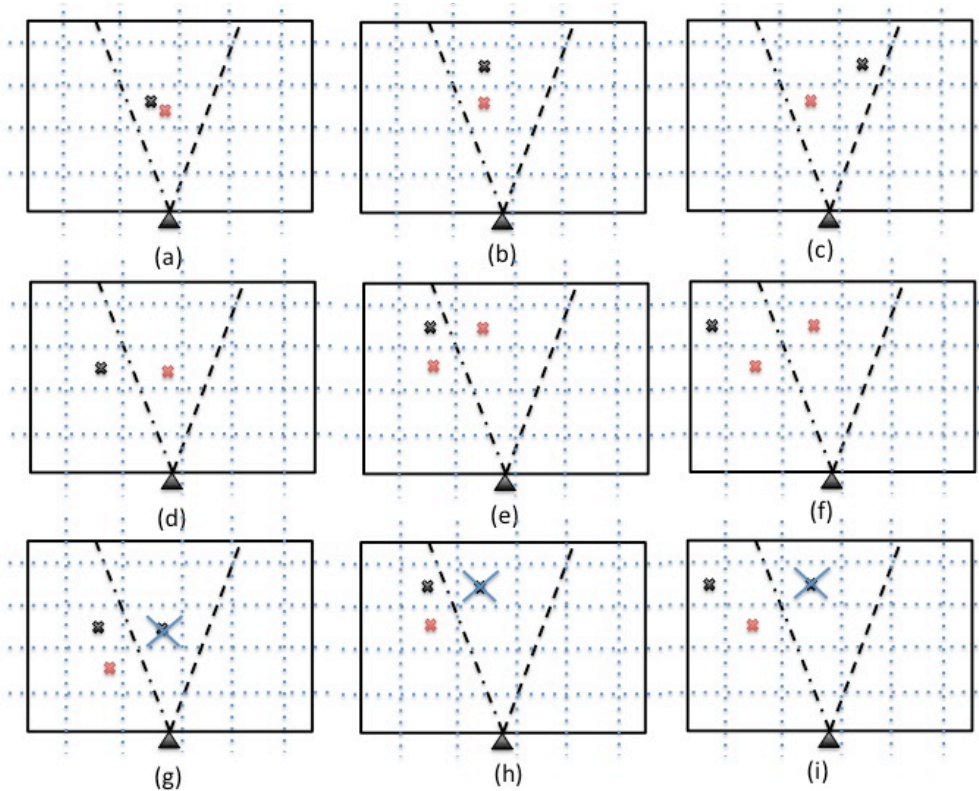


Figure 32 : We identify instances with our grid pattern and create the corresponding distance information.

5.7 Evaluation

Image to audio mapping is a tricky and challenging problem [69]. The visually rich information is reduced to a very low bandwidth and presented to the VI users to use. The goal of this study is two fold. Firstly, we would like to test the principles derived from our case study can indeed be broadly applied. The success of the mapping from a given floor plan to a verbal description can be determined by the perception and interpretation of the output by the end users. The measures used to

record the user perception of the problem are from the qualitative feedback on the user experience and usability evaluation. The measure used to record user interpretation is by asking the subjects recreate the floor plan as described by the verbal description. Secondly, we evaluate the end-to-end system and the interface. The experimental protocol used to observe the interpretation of our results by the end users is described below.

Ten individuals with visual impairment were involved in this experiment to evaluate our system. We have 18 participants in this stage of the experiments, 9 of whom were born blind and the other 8 participants lost their vision during childhood. All of them are cane users and 4 of them have a guide dog for their everyday navigation help. All the participants are avid Internet users. We have 6 low vision participants. The average age of the participants is 35.2 with the youngest participants being 21 years and the oldest being 54 years. The four participants who took part in our initial case study also took part in the experiments. We included them in our experiments because our evaluation is activity based. The efficiency with which they can complete the task will solely depend on if our verbal description delivers the message across verses their bias towards a certain way of description. We do not use sighted blind folded participants in our study because our verbal description system is purely based on the method of description based on our case study. Also, the criticism of the description obtained from the sighted volunteer, points to the fact that something rated highly by sighted people may not be suited for people with visual impairment.

Table 5 : Background details of our participants.

Participant id	Visual condition	Age, Gender	Occupation	Additional info
1	Low vision. Progressively lost vision during teens.	23, F	Student	Jaws user. Relies on residual vision during daytime.
2	Adventitiously blind. Lost vision when 8 years old.	21, F	Teacher for special need students	Jaws user. Uses cane and guiding dog.
3	Congenitally blind.	43, F	Orientation and Mobility Instructor	Jaws user. Uses a cane. Proficient with technology.
4	Completely blind. Lost vision in her teens.	21, F	Graduate Student, Biochemistry.	Jaws user. Uses a cane.
5	Congenitally blind.	49, M	Online English tutor	Cane user. Uses echo locator techniques for navigation. Very tech savvy.
6	Congenitally blind.	32, F	Braille proof reader	Cane user.
7	Lost vision at 32. Low vision.	35, F	Graduate student	Cane user. Relies on her residual vision a lot.
8	Congenitally blind.	46, M	Braille Proof reader.	Cane user. Very active and well versed in accessibility technologies.
9	Congenitally blind.	25, M	Freelance journalist	Cane user.
10	Low vision. Congenitally blind.	21, M	Intern in Senators office.	Athlete. Uses a cane.
11	Low vision. Congenitally	32, F	Unemployed.	Cane user.

	blind.			
12	Lost vision during an accident.	46, M	Was a Fedex driver before the accident, currently unemployed	Cane and guide dog user.
13	Low vision. Congenitally blind.	24, F	Orientation and Mobility instructor	Works with blind children. Cane user.
14	Low vision. Congenitally blind.	49, M	Unemployed.	Jaws reader.
15	Born with Melasma	52, M	Retired Office Assistant	Can rely on residual vision during the day. ZoomText user.
16	Completely blind. Lost vision at age 3.	54, F	Office Administrator	Jaws reader. iphone user. Leads a very independent and active life. Travels with a guide dog and a cane.
17	Low vision. Has been losing vision since she was 16.	25, F	Student.	ZoomTech user. Travels a lot.
18	Adventitiously Blind. Lost vision at 18.	25, M	Student	Extremely tech savy. Programmer. Travels with a Guide dog.

Our system is a web-based application, which is currently being hosted on a webserver and can be opened from a browser and used in conjunction with a screen reader. We use a windows machine, (i5 processor, 8GB RAM, 64 bit operating system) with Jaws 6.0 to test our system. The user was asked to sit comfortably in front of the PC. We opened the webpage and started the screen reader prior to the user starting the experiment. For the testing of our system, we used a puzzle like

reconstruction exercise to test the participant's understanding of our system and the proposed verbal description. To test the participants understanding of shape, we used long swell paper cutouts containing raised dots on it. We provide the user with multiple strips and ask the user to recreate the outline of the shape. We next provide the user with small triangular blocks to place along the rectangle to indicate the entrances. We skip these two steps for users who have never worked with tactile representations of shape. We created braille labels on swell paper with the names of the landmarks printed on them. The user is provided with a space on the desk adjacent to the computer reading out the verbal description, and is asked to recreate the building layout using these labels. For low vision users who do not read braille, we have used large print labels instead of braille labels. An overview of the system in use and the testing materials can be seen in Figure 33.



Figure 33 : The materials used in the experiments.

For the sake of practicality and usability, we confined our experiments to floor plans of libraries. Our approach can be extended to other building layouts as well. We made a list of local public libraries with an intention of providing information that they can actually use in planning a future trip. There is no underlying structure to these maps and they come various shapes, sizes and difficulty level. Difficulty levels are determined by the layout of the building, number of rooms in the floor, number of entrances to the floor, and the design tools used for the map when created. A restriction we placed on our dataset is to use ‘main level’ maps since they contain an entrance and provide most essential information for a first time visitor. Each user was asked to test the system on four of the five libraries. These libraries were chosen on the criterion of time efficiency w.r.t the web crawler, complexity of the building and with consideration of number of entrances. We made sure each user tested the one building that had two entrances. We used three maps for the sake of reconstruction purpose. We used one additional floor plan for a listening exercise towards the end of the experiment.

We designed an experiment to test the assumption that our system aids an individual with visual disability in the formation of a cognitive map. The experimental procedure consists of three phases. In the introductory phase, we collect personal information about the participant and explained the general nature of the experiment. The second phase pertains to the evaluation of the correlation between our generated verbal description and the original map. We achieved this by asking the participant to recreate the map, after listening to the descriptions. They were told to listen to the description carefully without emphasis on remembering any of it. We asked them to concentrate on the terminology used and the information

presented. After listening to the entire description once, they are asked to go over it line by line. Only 10 of the 17 participants indicated that they could read braille. We created braille, English alphabet and audio labels and the participants could choose the medium they want to work with. As the participant listened to the description line by line, we handed out the appropriate tactile cutout, to save them the effort. A sample of these labels can be seen in Figure 30. After the completion of this exercise with the three maps, the participant was asked some questions for qualitative feedback. The last task was to listen to the description and answer some question about the layout verbally, without doing reconstruction exercise. We added this task to assess if the participants do find such a description useful without performing a reconstruction exercise using tactile cutouts. They were asked questions pertaining to flipping the map around and a basic question regarding the relative position of two landmarks in the map. In addition to the above mentioned activities, we recorded the time taken, number of times they asked for assistance and number of times they listened to the description before finishing the task. At the end of the experiment, the users were asked to take a short survey to record feedback.

The aim of this system was not to have the users recreate the floor map with infinite precision but to aide to provide an overall understanding. We wanted the users to firstly understand the overall building shape. This was tested out, by making the users recreate the shape using thin strips of Lego like paper cutouts with raised dots on them. Some of the users were not familiar with shapes in general; we skipped this step for those users. As long as the users got the shape right, we counted this task as a success. Next, we tried to convey the location of the entrance with respect to the building. The users were asked to locate the entrance

on the building outline they created by placing markers along the building outline. We realize it is unfair to measure the accuracy of this task based on exact distance, since its hard to work with a 24" by 36" board and keep track of spatial distance, for an individual with visual impairment. As long as the users could do so with approximate proximity, the task was deemed successful. We calculate the success rate in terms of overall and based on the demographic. The next task in the experiment is to listen to the verbal description and create a map capturing the relative positions of the landmarks. The input floor plans come with no scale information or cardinal direction information. It cannot be expected from the participants to get the distance very accurately. But as long as each landmark is located in the right direction, we counted the task as a success. We also photograph the overall recreated image to compare with the original. We have a skelitized version of the images shown in Figure 31 to ensure better comparison.

5.7.1 Results

The system comprises of the Browser Interface, a Landmark Generation Module and a Verbal Description Module. Because the main focus of this article is integration of the verbal description into the prototype system, we present a detailed analysis of the user feedback on it, in the next subsection. The users were asked to evaluate the overall system by taking the Questionnaire for User Interaction Satisfaction (QUIS) developed by Shneiderman (<http://www.lap.umd.edu/quis>). The system scored 7.9 on an average for overall reactions to the system on a scale of 1 to 9, with 9 being satisfying. We had full scores on accessibility, sequence of screens,

screen layout and learning the system. The user of terminology throughout the system scored an average of 8 on a scale of 1 to 9, with 9 being consistent.

The first part of the information presented to the users is the shape information of the building and the location of the entrance wrt to the shape. The participants used the tactile strips provided to provide a representation of their understanding of the building geometry. Almost all the participants created geometrically accurate maps. We had three bad cases though out the entire experiment. We had one participant (figure 34 (f)) who used a square instead of rectangle. But later when asked about the mix up, said they hadn't paid enough attention and was eager to do the exercise. Another participant mixed up the information given for length and width. She said it was confusing. We had one participant (figure 34 (e)) mark the entrance in the wrong location. When asked about it, he said the terminology was not clear enough. We had some participants who were not familiar with the terminology of 'length' and 'width'. The experimenter gave some introduction to the shape terminology used. One participant opted to not do this exercise since he was not familiar with shape information and said our introduction hasn't helped. For the sake of completeness, we include some results in this section. Shown in figure 34 and 35 are some reconstructed shapes. We show some good as well as the bad cases for two of the five maps used.

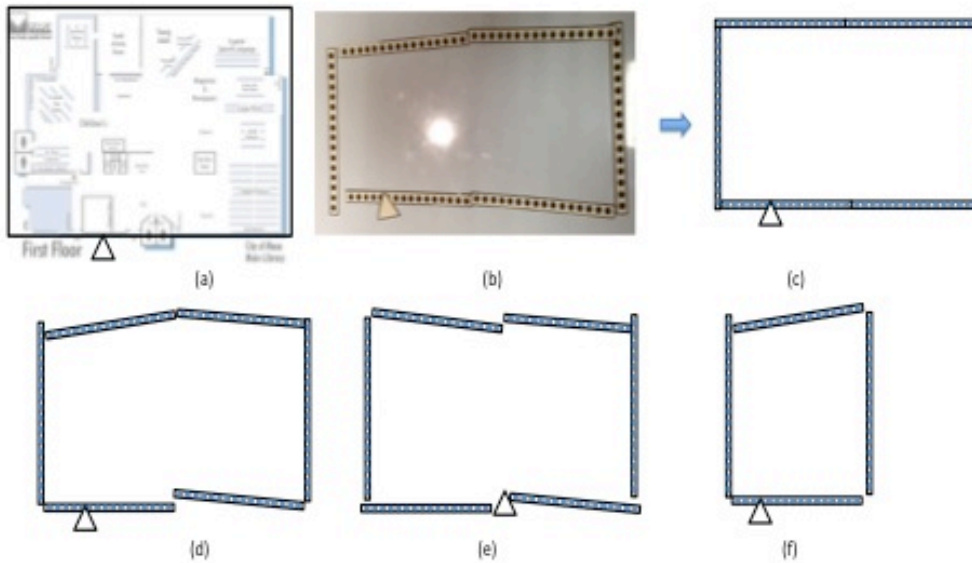


Figure 34 : A selection of results from a subset of users.

User Comments: Multiple participants mentioned the fact that when they read the description for the first time, they hadn't paid a lot of attention to the shape information. The exercise used for testing has helped with placing the shape information together with the landmark information presented in the description. Fourteen out of the eighteen participants thought shape information was important in the verbal description. They made comments like "now we know the side containing the entrance is the longer side", "this is also useful for the younger blind children", etc. The other three of the four participants thought it was good to have it available in the description but it did not contribute to the value. One of the participant mentioned that she can do without it. She was the participant who skipped this exercise. Eight of the eighteen participants heard the shape description multiple times, (on an average of two times,) before they did the experiment. The rest of the participants performed the exercise after listening to the description once.

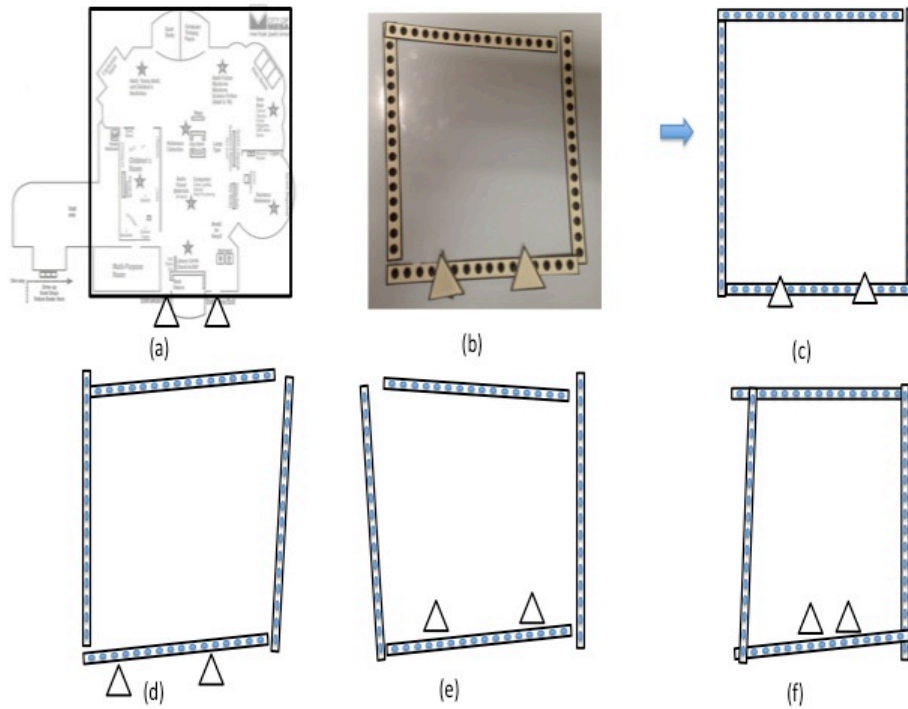


Figure 35 : Some sample reconstructed shapes using our system.

Participants were asked to listen to/read the verbal description and reconstruct the map, as they went through the description. For the sake of convenience, the experimenter handed the appropriate cutout after each sentence read. Figure 36 shows a sample reconstruction of a map by six users.



Figure 36 : The user recreated maps from cutouts and verbal description.

Quantitative Evaluation: The correlation between the original map and the reconstructed map can be evaluated by using the correlation coefficient calculated using Bravais-Pearson correlation coefficient. All the participants showed a correlation coefficient greater than 0.51, which indicates better than average reconstruction in terms of exact locations. The highest correlation coefficient obtained by a single user is 0.85. The average correlation coefficient for all the users is 0.62. This is not a high score in terms of reconstruction, but we need to take into consideration that the task of exact reconstruction is virtually impossible for the audience in question using a verbal description alone sans the distance information. We used a second metric to evaluate the positioning of the labels by the participants, without consideration of the exact distance. We used an 8-neighbor accuracy to assess the accuracy of placement. For each landmark, we recorded the accuracy in

directional location of its neighbors in all directions. All the participants showed an accuracy of more than 71.6 percent with an average of 89.3 percent. The highest accuracy as recorded per user was cent percent, indicating that the verbal description was easy to understand and work with. All the user scores progressively increased as they reconstructed more maps.

A concern we had when designing the experiments was the fact that the evaluation was based on asking the participants to physically reconstruct the map. In real time, the user will not have the materials to do this exercise. He would have to listen to the description and use his understanding to create his cognitive map. To test the effectiveness of the description sans the exercise, all the participants were given a fourth map, where they had to listen to the description and answer a question regarding the location of a landmark with respect to another. The users had to mentally flip the map to answer the question. The map picked for this exercise was constant for all the participants. The participants were allowed to listen to the description after the question was asked. All the participants got the relative location right. Some participants mentioned that this exercise may be hard for young kids who are blind and having the kids do the physical mapping is a good way to help them learn a building layout.

We tested three building layouts with each of the participants. We made sure that the user always starts off with the easiest layout first. In spite of this, we noticed that the users' comfort and ease of completing the task increased as we progressed through the maps. The users finished the tasks faster for the second and third maps when compared to the first. When asked about what contributed to this effect, they mentioned the fact that they got used to the terminology. We provided no

additional training before the start of the experiment. And the users were able to pick up the language and translate it into a physical map with minimal effort as we progressed into the experiment. This shows that our system can be deployed without a need to train the user.

One of the five maps introduced in the experiment contains the following scenario: Emergency exit is located in the middle of the building, and not necessarily towards the edge of the building. Nine of the participants immediately placed the exit towards the edge of the magnetic board using their intuition, and got confused when they heard about locations that lay “ahead” of this emergency exit. Three of the nine participants started reconstructing everything else flipped backwards towards the entrance from this point on. This was the primary reason for lowered rate of accuracy. One of the participants remarked that the system got the description wrong since our verbal description pointed to locations beyond the emergency exit. The rest of the nine participants realized the error and moved everything around to accommodate the rest of the landmarks. This also contributed to the reduced accuracy in the placement. This incident points to the fact that prior experience matters a lot when processing new information when it comes to navigation.

Another observation we realized during the experiment concerns distance. The system provided some information about distance like “immediate left”, “little further”, etc. The users would take notice of the distance, but this did not translate well on to the magnetic board. We heard they say phrases like “oh, this is really close to XYZ landmark” but the relative distance on the board when compared to other distances would almost be constant. This made us to believe that distance

recreation is harder to do in the absence of vision. But surprisingly completely blind participants made a better effort to follow the distance cues. Looking at the overall effort from the participants, they were more concerned with getting the relative positions accurate. Another general observation that is prevalent through the entire experiment is fact that it is not an easy task for an individual with visual impairment to assess distances on the magnetic board provided. Almost all the participants thought the tactile strips given to them for reconstructing the shape were of unequal length. The only distance information that was readily used was the words “immediate” left/right. The act of placing the entrance at the center of the magnetic board for the second part of the experiment was a hard exercise since it’s hard to assess the center just by using touch. The participants mentioned that the exercise was ‘fun’ in general, but the difficulty in placing the labels in the accurate locations was evident.

A limitation of the testing equipment is the use the rectangular tactile cutouts containing the labels of the landmarks inside the building. Some users wondered if they should place the rectangular labels sideways or straight up. The orientation effected the reconstruction to some extent, especially when considering accuracy in terms of distance. Some users asked if the rooms were rectangular based on the shape of the tactile labels. Using equal sized cutouts required a bigger magnetic board that was no longer “at arms reach” making it harder for the participant to work with. Hence, we went ahead with the smallest possible individual cutout sizes.

One of the participants had suggested an alternate scheme to generate the description. She wanted us to describe all the locations in the center of the building

and then describe the left and right segments with respect to what has already been described. She indicated that such a scheme would make it easier to do the reconstruction puzzle we made her perform. But when asked what she would prefer in the absence of the reconstruction puzzle and solely based on listening to the description, she felt the current description scheme to be more apt.

The users rated the verbal description on a scale of 1 to 5, with 5 being the hardest. The overall rating for the system averaged out for all maps and users is 2.1. We noticed that the users rated their first maps the highest on the scale and the rate of difficulty went down as the users progressed through their second and third maps. We had several users mark zero as the scale of difficulty on their third map. The participants indicated that it took them the first two maps to get acquainted with the language and pay closer attention to the distance information provided.

The participants thought this system is useful for kids with dyslexia and individuals with visual impairment. All the participants indicated that they would recommend this system to someone with visual impairment. We asked all the users about the drawbacks of the system. A participant indicated that the lack of cardinal directions limits the usefulness of the system. He wanted the location of the entrance with respect to the cardinal directions. He further wanted us to aggregate information about the building in terms of the contact information like the phone number, landmarks and location of the building from Google maps to make it a one stop visit website before he travels to some place new. A few participants suggested having an additional interface to choose the level of details and the entrance of their choice.

Low Vision Vs Complete loss of vision : In our experiments that test cognitive mapping, we ask the participants to reconstruct the map, as they hear it on a magnetic board. The participants hear terms like “to the immediate right”, “little further”, “further ahead”, etc. There exist two components to realizing this information for the sake of the experiment. Firstly, the users need to pay attention to the distance information. Secondly, the users need to be able to use the distance information to recreate the map. It is our understanding from looking at the reconstructed maps that the completely blind users paid more attention to the terminology and tried to use the distance information in the reconstruction. On the other hand, a lot of low vision users discarded the distance information, used a small space on the magnetic board concentrated around the entrance to do the reconstruction. End of experiment survey records state that all the users noticed the distance information. But the completely blind users could translate this onto the board more effectively. This could be because they paid more attention to the description or due to the fact that due to the lack of residual vision, they are used to paying more attention to subtle details in general.

Gender : Through the experiments, we note that the female participants were more anxious to do the reconstruction exercise perfectly when compared to their male counterparts. They asked us questions on their accuracy once they were done with the experiment. They tried to analyze the reason why they got something wrong by going through the description and suggesting possible reasons. Most male participants showed a more nonchalant attitude towards the experiment. They concentrated more of visualizing the layout and treated the exercise as a hindrance to the listening exercise. They also commented on how positions inside the building

made logical sense. For example, one of the buildings had book check out near the entrance. They commented on such observations (“this makes sense”, ”this is a good layout for the library”). For the sake of the experiment, we reduced the number of landmarks presented to the participant. This resulted in some blank spaces in the reconstructed map. The male participants repeatedly asked us questions on what lay in these blank spaces. Two female participants (id: 16 & 13) asked us similar questions, but they represent are a very small fraction owing to their active lifestyle and profession.

Age : Our participant ages ranged from 21 to 54 years. We couldn't conclusively find any differences in the accuracies of map reconstruction based on age. Going through the experiment, we found that the mid age group (30-40 years) participants were the group most critical of the description. They readily commented upon any lines in the verbal description that they found confusing. They readily asked for more information like location of parking lots, etc. The younger group could finish the experiment in the least amount of time and could go through the description quickly. We couldn't obtain any suggestions and negative feedback from this age group. This could be because they are in their prime years and have enough experience to work with the available information without further edits. The older generation was extremely appreciative of the information and did not have much criticism or suggestions when it came to the system. This could be due to the fact that they have managed to navigate most of their lives with little information and without a lot of technology.

Age of onset of visual impairment : We categorized our participants into two groups: the congenitally blind and the ones who lost their vision later in their life. It

was our assumption that individuals who are born completely blind have more trouble with visualizing information. But this turned out to be false during the study. Both groups of participants performed equally well during the experiment with the reconstruction. During the final questionnaire, multiple participants mentioned the fact that visualizing a space using verbal description alone may not be possible for all blind users, especially the ones who were born without vision. We haven't encountered any in our small participant pool.

Everyday mobility : Our participant pool included people with a wide range of occupations and various levels of everyday activity. The director of the disability services for university is known to manage all her activities on her own and has experience visiting a lot of new locations. The students who participated in our study walk the campus everyday. The O&M trainers are on their feet a lot. We have some participants who are older/retired/unemployed who don't lead as active a lifestyle as their employed counter parts. We could notice a stark contrast in the processing time between the time the participant listens to the description and the reconstruction exercise. The experiment was not a timed exercise, but the difference in the time taken to finish it was notable for this demographic. This could be due to the exposure to tools, more visits to unfamiliar locations and to asking and following directions in general. But this did not affect the accuracy of reconstruction.

Computer literacy : We had two participants who are currently enrolled in a class that teaches computer skills. They were learning how to use windows environment, screen readers, etc. They do not use email and we had to use telephones for communication. We gave them a small overview on how to use the system and they could catch on easily. This shows that the system is easy to learn

and use. They did not report any problems with using the interface and didn't need any extra assistance.

Orientation and Mobility Instructors : Two of our participants work as orientation and mobility instructors at a local organization for blind children and are visually impaired (refer table 1). O&M training is highly recommended for individuals with visual impairment and helps with valuable life skills. One of the modules is about navigating indoors. The module is aimed at helping the individuals navigate an unfamiliar space using shorelines, audio cues, keeping track of directions, etc. The instructors make good critiques for the proposed system. In addition to the tasks performed by all the participants, we have included additional questionnaire to get their feedback. The questionnaire concerned the grammar used to create the verbal description, the terminology used for indicating direction and distance, and they were asked to comment upon any sentence in each of the descriptions, which could be confusing to the audience. We report an additional metric: percentage of total sentences marked as ambiguous by the instructors. Out of the 143 lines of descriptions pertaining to the 6 maps evaluated by the two instructors combined, we have 6 lines flagged as ambiguous by the instructors. One of the flag was regarding encountering Emergency Exit in the middle of the building. The overall feedback towards the system has been extremely encouraging.

Users from the initial study : Four participants from the initial study participated in the above-mentioned experiments. The language and grammar was derived from the descriptions they wrote for a building they are very familiar with. We heard comments like "oh, I like how it stays to the immediate right", "straight ahead, that's easy to follow", from these participants. One of the four participants

used the following scheme during the case study: she first described all the locations encountered walking straight into the building followed by left and right location. This participant mentioned this again as her preference during the experiment.

Zoom-tech users (with residual vision) : Two of the participants, technically classified as low vision users, had some residual vision. They used ZoomTech (<http://zoomgroup.com/>), a screen magnifier instead of the screen reader. Since our system provides the original image of the map along with the verbal description, they could access a zoomed version of the image. After each reconstruction experiment, we asked the user to go through the image. Firstly, we observed that in spite of the availability of the image, the participants had a hard time scrolling through the enlarged image. Both the users indicated that they preferred the verbal description to the image. This was evident from the trouble they faced when trying to go through a zoomed image that lost its sharpness. The participants were asked to comment upon the descriptions going through the image. One of the two participants faced a lot of trouble scrolling through the image that we called off this task. The other participant could not assess distance since scrolling when zoomed in was not an easy to do task. He gave us very positive feedback on what he thought about the correlation between the image and the description that we generated.

5.7.2 Discussion

A new web based assistive technology was designed to aid individuals with visual impairment “read” floor layouts of unfamiliar public buildings. Its effectiveness was tested by asking the participants to replicate the layout and by answering questions regarding spatial relations of landmarks. Evaluation using

correlation coefficient and accuracy of locating neighbors indicates that most participants were able to realize a cognitive map using our system. This system might be an effective aid for individuals with visual impairment to learn the layout of a new location prior to visiting it. The fact that most public buildings are required by law to publish a map online makes this tool ready to use in real time compared to the existing contemporary works that require special formats and human intervention to provide the required information.

5.7.3 Use Cases

To better understand the prototype system we designed and built, we present some use cases. A student at Arizona State University wants to visit the schools main library. He opens our interface on his laptop and puts in the keywords “library ASU University” into the search box. Alternately, if he is familiar with the website for the library, he would say put in “lib.xyz.edu” into the search box. Our system will search the website, download the floor map if available and save a text file on the desktop of the user containing the generated instructions. The user listens to this file using synthetic speech software that is installed on his computer. The second scenario we describe is something we have noticed working with individuals with visual impairment and wish to alleviate. A blind student regularly visits a Disability Access Consultant in a designated building. He is very familiar with getting to the specific office from a designated entrance to the building. We realize that many students follow the exact same route in and out, and seldom have any idea of other useful service providing facilities in this building because they haven’t explored the area. Having a verbal description helps them form a spatial map in terms of what

they already know, make it easier to locate new rooms and perhaps make them bolder to explore around a familiar location. An integral part of Orientation and Mobility Training (OMT) for children is using a sighted person to aid in travel. Our system can be used to train children to look for information about the place before going on location, by giving the description of a new place thereby encouraging children to try exploring new places independently.

5.7.4 Limitations

We acknowledge that there is not universal idea way to generate a verbal description for every sort of building layout and we do have some limitations to our approach. Firstly, keyword based URL search could fail if the site architect fails to use at least one of the relevant keywords in the URL/filename or labels other images with these keywords. Secondly, our system produces the above presented results for maps of medium-good resolution. This is a constraint that ensures the robustness of the OCR module. We could not detect a small percentage of some important landmarks, when the OCR module couldn't not handle the font variations and the cases with lower resolution. Thirdly, most floor plans downloaded from the web lack orientation and distance scale information. Under these circumstances, we can only present relative information effectively. Lastly, impaired vision could be onset by old age and other factors, which affect the physical and cognitive abilities of the individual. Younger individuals have good learning abilities that can be seen from our experiments. We could not test on significantly older blind users to test our prototype. A constraint we used to pick our dataset is the format used to present the floor plans. Many establishments have multiple floor plans published in one

image/PDF file. These add an additional task to our approach to extracting each floor map out. We currently do not handle these cases. On that note, the system can be improved if the site architects used higher resolution images and uses more standardized architectural symbols. Hopefully, presence of a system like this would motivate more establishments to publish their floor plans and follow a standard format.

5.8 Using Beacon Technology

So far, we developed a system that devised a method to present information from a floor plan to an individual with visual impairment. Through this process, we also derived a few paradigms of how to present information on indoor locations. We explore this idea further by considering a real time application that uses the the Beacon technology and the verbal description module. iBeacons are low energy proximity sensing device that transmits a universally unique identifier that can be interpreted by a compatible application. These devices can trigger location based prompts and have been used as contextual commercial applications. Every beacon ID is 20 characters long and is divided into three parts. A 16 byte UUID is the master ID that is first in the hierarchy when distinguishing between beacons. Typically, when installing the beacons, the entire institution gets one ID. The second ID in the hierarchy is a 2 byte ID. The major can be used to distinguish between buildings or floors or departments. Next in the hierarchy is a 2 byte minor ID that is unique for each beacon. We have used beacons built by Estimote (www.estimote.com) and used their API for developing a prototype.

5.8.1 Case Study

The first step towards building this by obtaining some feedback from the Orientation and Mobility trainers. We walked around the building with them and they gave us some wisdom as to techniques we should pay attention to. The responses helped us learn some new perspectives. Firstly, when giving directions, using another object that they cannot see as a reference is not very helpful. For example, “take a right after you pass by the printer”. This would involve them using their cane and hands to figure out where the printer is positioned before taking the next step. But if the object was noisy or has a scent, it would be helpful. Example : “near the coffee machine”, “near the escalator”, etc. We found out that a surprisingly few object fit this criterion. Secondly, they noted that hallways have multiple pathways on either sides. And instead of using phrases like “walk down the path way and take a right”, we should try to use some metric that they can measure like “3 doors down to the right”/”15 feet along the hallway”. Thirdly, when they train individuals, they encourage the individual to explore the region instead of memorizing a route. This exercise involves walking with the blind individual and giving him useful information about the surroundings as they walk. They recommended us to try emulating this principle when building the app. For the first prototype, we decided to implement a system that helps the user explore a new indoor space.

5.8.2 Calibration

The first step in using our system is to place the beacons. We are using Estimote’s Indoor SDK [70] for the calibration. This is done in two stages. In the

first stage, the user walks along the edges of the room. The application uses a combination of the compass, gyroscope and GPS coordinates to assess the shape of the room. Next is beacon placement prompts based on the size and shape of the room. The SDK uses uniform placement to make the initial suggestions. But this step is not optimized for signal strength. We add a further step to this process by asking the user to walk along the placed beacons one more time. We record the signal strength from the closest beacons along the trajectory that the user walks. We look for weak hot spots and suggest possible locations for better signal distribution.

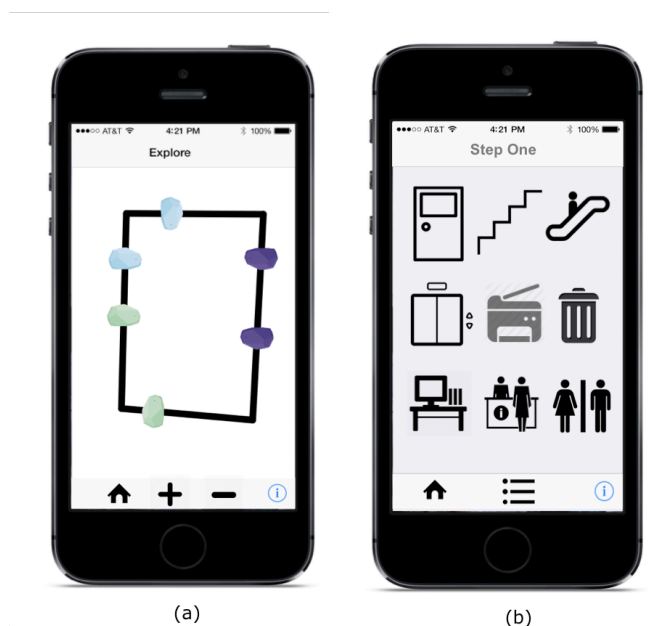


Figure 37 : (a) shows a sample calibrated space. The beacon placement on the wall is illustrated. (b) shows a method of input the information inside the space.

5.8.3 Information Gathering

Furniture, obstacles like columns/podiums, cubicles, conclaves with entrances, rooms, walkways, hallways, reception counter, coffee area, We built a tagging system that a user who is sighted and has administrative privileges can use to enter the information. Placing the beacons is a preliminary step but the spatial

information about the landmarks in the building needs to be fed into the system. We used the Indoor Location SDK to generate a global coordinate reference system for the space in question. The user can then walk around and tag each landmark. The system collects the coordinates of the landmark and stores it in the database with its textual information given by the user. Instead of giving the user a no rule unstructured format for this input, we create a custom input prompt.

Giving the administrator an open ended input system meant that our system should be able to understand the words used for tags and use it. If someone tags a 'card board box' laying in the corner of the room or a 'chair' in the middle of the room, would we have to resort to using these obscure landmarks the way they were tagged. This presented a problem with information collection. Instead, we decided to categorize architectural elements as a menu to pick from. The user picks a category and then enters the tag. This can be seen in Figure a. The first screen shows a menu of commonly used architectural symbols that exist inside the buildings. We do not have a comprehensive list as yet, but have 15 items for the user to choose from. Once the user chooses one item from the menu, we assist the user to collect information required to properly mark this landmark. For example, if the user chooses to enter a room into our system, he is given two options. He could mark the entrance. Or he could mark the corners of the room. Based on the information entered, we obtain accurate information to assist a user. We also allow the user to enter custom tags that do not fall into any of our 15 categories. These tags are processed differently.

5.8.4 Tired Information Delivery

We present information in two stages. First is the overview of the entire space. Second is information about the surroundings as the user walks around. This is further classified based on importance. We collect information in one of the 15 categories from the admin. This information gets higher priority over other kinds of tags that the admin may choose to user. We also collect information regarding custom messages to be delivered to the users. This may be a change in the room number, alert about an obstacle, etc. This information has the highest priority and gets delivered when the user enters proximity.

5.8.5 What's around me?

Using the calibrated beacons and the inbuilt compass, it is possible to accurately measure the position and direction of the user. Based on this, we deliver information about the surroundings. We worked with a 70 square meter space and used a threshold of 2 meters as a trigger to set off the proximity alarm. We used the ego-centric method to deliver this information [59]. This information was presented as audio and as text on the screen.

5.8.6 Experiments

The claim we make is that an individual who uses our application the way its intended helps create awareness of the space around him and helps him form a cognitive map of the surroundings. We tested this claim by enlisting the user to perform a navigational task and a reconstruction task. We had 3 participants evaluate our system. We had 2 O&M instructors test each stage of the application. One of the O&M instructor is partially blind resulting in a different perspective on

the feedback obtained. We had one individual with visual impairment come in and test our system once calibrated. We had one sighted person in the role of the administrator for entering the “crowd sourced” information.

5.8.6.1 Verbal Description

O&M manuals and literature proves that providing the user with an overall description indeed aides with formation of a cognitive map. We did not set out to reprove this facet of our system. We let the user listen to the overview as many times as he needed to. We evaluate this part based on quality of the information presented. The users were asked to rate the description based on clarity, amount of information presented, perceived usefulness of information presented. The average rating on usefulness of the description is 4/5.

5.8.6.2 What’s around me

The users were asked to walk around the space at their own pace. They were allowed to revisit the landmarks as needed. They get to hear the details of landmarks around them based on their position and direction. This feature is present in every navigational application that has been designed for individuals with VI. We did not set out to test if this feature is indeed useful but tested the quality of information being delivered. The user was asked to rate the description based on accuracy (score = 3/5), clarity (score =5/5), amount of information presented (score = 4/5), perceived usefulness of information (score = 4/5) presented.

5.8.6.3 Walking assignment

Once the user spent some time walking around the space and using the app, they were taken back to the entrance. He was asked to listen to the verbal

description again and then asked certain questions. The questions ranged from closest landmark to potential direction of landmarks. And then, they were given a task of finding a certain landmark that the user passed by during the training stage. This wasn't a timed exercise. We asked the user to think out loud as they completed the exercise.

5.9 Conclusion

We have used a dictionary of keywords to select a subset of landmarks from the libraries we tested our system with. This was done to reduce the number of landmarks the participant needs to reconstruct for the sake of the experiment to a maximum of 20 landmarks. But we acknowledge that a building can have anything upward of 10 landmarks and can go up to 100+ landmarks. We explained this to the participant and presented them with the following interface options for the future. The first interface would have list of all the landmarks and an adjacent check box. We then refine our description to include landmarks that surround the point of interest. This option received the most number of votes. Some participants suggested we give them an option to choose "Show important/Show all". The third interface suggestion was a tiered representation. Give the user a base layer description. And build on the finer details around a point of interest. By show of preference, we do not have a clear winner on which of these the users might prefer. But in our future work, we want to explore building another layer to the current interface where the user gets to exercise more controls on what sort of information he receives from the system. Although our studies show that having angle based approach or allocentric approach were not preferred, we think one solution fits all in

the day and age of interfaces makes for bad design. We want the users to choose what sort of verbal description they would like to hear and choose the one that they feel comfortable using. We have not done this currently because our work focuses more the building of fundamental components and knowledge gathering. We have not included additional interface experiments in this work to keep the emphasis on effective way to generate and deliver verbal descriptions for indoor floor plans. We extended this work to using beacons in indoor locations and calibration support on iOS. The preliminary experiments testing this application indicate a promising lead. But a lot more development work needs to be undertaken to increase the accuracy of the indoor positioning system. This work is an ongoing effort in my research group.

CHAPTER 6

CONCLUSION

Until the last decade, solutions to aide individuals were pure research prototypes or standalone devices that require an investment. In this dissertation, I have explored accessibility using universally available everyday devices. The possibility of developing accessible application on the web and mobile led me to explore developing solutions for commonly faced problems in the everyday life of an individual with visual impairment. Specifically, I looked into content aggregation and smart interfaces. Firstly, we looked into accessibility of static webpages versus web 2.0. W3C guidelines provide sufficient help for creating HTML pages that can be read by screen readers. But for intelligent ‘news feed’ type of websites like Amazon, Facebook, Youtube, etc, making buttons or images readable does not solve the problem. My work looked into how architecture of the website and layout makes an impact. We also proposed a method to develop dynamic content that brings the useful content to the user using Hierarchical Dirichlet Processes.

Secondly, we looked into finding methods to aide with navigation. Tactile based maps have been around for decades. We looked into how we could present the vital information without using special printers and material that requires to be carried around. iOS has a built in accessibility mode making it the device of choice for all of my applications. It is a common consensus from our case studies that the information available on the web catering to sighted individuals is not generally sufficient to build applications that can assist blind individuals. A big chunk of my work is to identify the missing information, devising methods to mine the needed information and present it in the most accessible way. We have employed the use of

web crawlers, crowdsourcing, trained volunteers and image understanding techniques to obtain the needed information. We put in methods to ensure quality control and recruited users to perform complicated tasks in real time. The results from our small group testing practices have indicated that it is a promising lead to pursue.

One of my proudest contribution is the identification of the O&M techniques from a 600-page manual that can be deployed by an application without the supervision of a sighted trainer. These techniques have been perfected over a century and contain a wealth of information in understanding the right way to present information. This work is by no means complete and there are many techniques in the manual that would require human-computer-interface designers to translate into usable applications. I would hope that my future work in the field of accessibility will continue this effort. Here is a summary of my findings:

1. Personalization is the key to designing systems for individuals with disabilities. From our experiments, we have met people with various abilities and cognitive abilities. Being fixated on the best method to present information is a first step towards building these systems. But letting the users personalize the information delivery is equally important. There is no one best way to present information for everyone.

2. Next generation systems should employ artificial intelligence to learn the user and adapt. Many a time, the users do not know the bounds of technology or can choose the settings will full understanding. The case studies I have used involved questionnaire and self-assessment. This step needs to be replaced by data collected over time.

3. Indoor navigation systems have not received enough attention and are equally important as systems for outdoor navigation. While egocentric and allocentric approaches work for lot of users, indoor navigation is much more complex and is very dependent on the environment and the user. The hybrid approach we proposed is a good start to develop a system.

I hope the systems of the future are more intelligent, tailored and user friendly. Finally, I would like to thank all the participants who took the time out to participate in our experiments and give us their feedback. I hope our research lay a foundation for future systems that aide the individuals with visual impairment.

REFERENCES

- [1] Section508. [Online]. <http://www.section508.gov/index.cfm?ID=3&FuseAction=Content>.
- [2] H., et al., Mankoff, "Is Your Web Page Accessible? A Comparative Study of Methods for Assessing Web Page Accessibility for the Blind,".
- [3] [Online]. www.w3.org/WAI/EO/Drafts/UCD/questions.html.
- [4] J., 2008. Howe, *Crowdsourcing: Why the power of the crowd is driving the future of business.*: Random House Publishers.
- [5] Charlene Li and Josh Bernoff., *Groundswell: Winning in a World Transformed by Social Technologies*.
- [6] R., Dabbish, L., Kraut, R., Postmes, T. Farzan, "Increasing Commitment to Online Communities by Designing for Social Presence.," in *CSCW*, 2011.
- [8] Y. W., Jordan, M. I, Beal, D. M. Teh, "Hierarchical Dirichlet Processes," Department of Statistics, UC Berkeley, 2004.
- [7] D. M., Ng, A. Y., Jordan, M. I. Blei, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 2003.
- [9] K., Banovic, N., Truong, K. N. Yatani, "SpaceSense: Representing Geographical Information to Visually Impaired People Using Spatial Tactile Feedback. ," in *CHI 2012*, pp. 415-424.
- [10] N.A., & Legge, G.E. Giudice, "Blind Navigation and the Role of Technology," in *Engineering handbook of smart technology for aging, disability and independance.*: John Wiley & Sons, 2008.
- [11] Y., Place, W., & Angeles, L. Lin, *Map-Enhanced UAV Image Sequence Registration and Synchronization of Multiple Image Sequences.*, 2007.
- [12] C., Stefanidis, A., Croitoru, A., & Agouris, P. Wang, "Map Registration of Image Sequences Using Linear Features. ," in *Photogrammetric Engineering & Remote Sensing*, 2008, pp. 74(1), 25–38.
- [13] B., & Flusser, J. Zitová, "Image registration methods: a survey," in *Image and Vision Computing*, 2003, pp. 21(11), 977–1000..
- [14] C., Knoblock, C. a., & Shahabi, C. Chen, "Automatically Conflating Road Vector Data with Orthoimagery.," in *GeoInformatica*, 2007, pp. 10(4), 495–530.

- [15] C., Stefanidis, A., Croitoru, A., & Agouris, P. Wang, "Map Registration of Image Sequences Using Linear Features. ," in *Photogrammetric Engineering & Remote Sensing*, 2008, pp. 74(1), 25–38.
- [16] M., Nationale, E., Ima, D., & Cedex, F. P. Roux, "Automatic registration of SPOT images and digitized maps," in *IMA*, 1996, pp. 46, 8–11.
- [18] Bing, and Baba C. Vemuri. Jian, "Robust point set registration using gaussian mixture models.," in *Pattern Analysis and Machine Intelligence*, 2011, pp. 1633-1645.
- [17] C., Knoblock, C. a., & Shahabi, C. Chen, "Automatically Conflating Road Vector Data with Orthoimagery. ," in *GeoInformatica*, 2007), pp. 10(4), 495–530.
- [19] [Online]. <http://maps.google.com/help/maps/floorplans/>
- [20] [Online]. http://www.huffingtonpost.com/2011/08/04/bing-maps-mobile-mall_n_918766.html
- [21] Tony Shu-Hsien Wang, "A recommendation for designing mobile pedestrian navigation system in university campuses," in *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*.
- [22] A., Riebeck, M., and Kawalek, J. Stark, "How to Design an Advanced Pedestrian Navigation System: Field Trial Results.," in *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 2007.
- [23] Purvis Ponder Everett Hill, *Orientation and Mobility Techniques*.
- [24] M., Gallagher, T., Li, B., Dempster, A. G., and Rizos, C. Tao, "Indoor/Outdoor On Campus Navigation System," in *International Conference on Indoor Positioning and Indoor Navigation*, 2012.
- [25] Bruce Blasch Richard Welsh, *Foundations of Orientation and Mobility*.: American Foundation for the Blind, 1980.
- [26] Ian Pitt Tracey J. Mehigan, "Harnessing wireless technologies for campus navigation by blind students and visitors," in *Proceedings of the 13th international conference on Computers Helping People with Special Needs*, Linz, 2012.

- [28] A.S., Moore, S.E., Ramachandran, B. Helal, "Drishti: An intergrated Navigation System for Visually Imparied and Disabled.," in *Fifth International Symposium on Wearable Computers*.
- [27] E., Medaglia, C.M., Serbanati, A., Ceipidor, U.B D'Atri, "A System to Aid Blind People in Mobility: A Usability Test and Results," in *Second International Symposium on Wearable Computers*.
- [29] Ang, L. M., Seng, k.P., Kong, J.H., Lo, R., Khor, M.Y Tee. A., "SmartGuide System to Assist Visually Impaired People in University Environment," in *Third International Convention on Rehabilitation Engineering & Assistive Technology*.
- [30] Parag Chandakkar, Peng Zhang, Baoxin Li Devi Paladugu, "Supporting Navigation of Outdoor Shopping Complexs for Visually-Impaired Users Through Multi-model Data Fusion," , 2013.
- [31] ROBERTS A. Perspectives from a Blind Point of View 2010. Returning to school.. [Online]. <http://www.blindgal.com/2010/01/returning-to-school.html>
- [32] R.G., MARSTON, J.R., LOOMIS, J.M., KLATZKY, R. L. GOLLEDGE, "Stated preferences for components of a personal guidance system for nonvisual navigation., ," *Journal of Visual Impairment and Blindness*, pp. 98, 135-147. , 2004.
- [33] B. Li, T. Hedgpeth, T. Haven Z. Wang, "Instant tactile audio map: enabling access to digital maps for people with visual impairment.," , 2009.
- [34] Z., LI, N., LI, B. WANG, "Fast and independent access to map directions for people who are blind.," *Interacting with Computers* , pp. 24, 91-106. , 2012.
- [35] D.A., MAGULURI, H., TIAN, Q., LI, B. PALADUGU, "Automated Description Generation for Indoor Floor Maps. ," in *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility.*, 2012.
- [36] J.M., FRANK L BARLOW, "Crossroads: Modern interactive intersection and accessible pedestrian signals. ," *Journal of Visual Impairment and Blindness*, pp. 99, 599-610.
- [38] C.-C., Knoblock, C. a., & Shahabi, C. Chen, "Automatically Conflating Road Vector Data with Orthoimagery," *GeoInformatica*, 2007.
- [37] J. GARDNER, "Access by blind students and professionals to mainstream math

- and science.," in *Computers Helping People with Special Needs 8th International Conference, ICCHP*, 2002., pp. , Vol.2398, 2002, pp. 502–7.
- [39] J., XIAO, Y., NOREIKIS, M., OU, Z., YLA, J.A. 2015. DONG, "iMoon: Using Smartphones for Image-based Indoor Navigation. ," in *ACM Conference on Embedded Networked Sensor Systems*.
- [40] Bing, and Baba C. Vemuri. Jian, "Robust point set registration using Guassian Mixture Models," in *Pattern Analysis and Machine Intelligence*, 2011.
- [41] W.J.M. LEVELT, "Language and Space," in *Perspective talking and ellipsis in spatial descriptions.*, 1996. , pp. 77-107.
- [42] G.H., KIM, T., LI, K.J. RYU, "Indoor Navigation Map for Visually Impaired People. ," in *Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness.*, 2014.
- [43] C., Stefanidis, A., Croitoru, A., & Agouris, P Wang, "Map Registration of Image Sequences Using Linear Features.," in *Photogrammetric Engineering & Remote Sensing* , 2008.
- [44] J. M., KLATZY, R. L. AND GOLLEDGE, R. G. LOOMIS, "Navigating without Vision: Basic and Applied Research. ," *In Optometry and Vision Science*, pp. 78, 5 (pp 282-289). , 2001.
- [45] R. M., STEA, D. DOWNS, *Image and Environment: cognitive mapping and spatial behaviour.*, 1973.
- [46] V. S. M. SENDEN, *Raum und Gestalt Auffassung bei Operarten Blindgeborenen vor und nach Cer Operation. Leipzig: Barth.*, 1932.
- [48] S. 1988. MILLAR, "Models of sensory deprivation: the nature / nurture dichotomy and spatial representation in the blind.," *International Journal of Behavioral Development*, pp. 11, 69-87.
- [47] J. J., GUTH, D. A. & HILL, E. W. REISER, "Sensitivity to perspective structure while walking without vision. ," *Perception* , pp. 15, 173-188. , 1986.
- [49] R.D. 1992. JACOBSON, "Spatial Cognition through tactile mapping. ," *Swansea Geographer.*, pp. 29, 79-88.
- [50] O. and MIODUSER, D. LAHAV, "Multisensory Virtual Environment for Supporting Blind Persons Acquisition of Spatial Cognitive Mapping – a Case Study. ," in *Proceedings of World Conference on Educational Multimedia*,

Hypermedia and Telecommunications. , 2001., pp. 1046-1051.

- [51] S., BLADES, M & SPENCER. S. UNGAR, "The construction of cognitive maps by children with visual impairments. ," in *The construction of cognitive maps*, 1996., pp. 247-73.
- [52] G.L. ALLEN, "From Knowledge to words to wayfinding: Issues in the production and comprehension of route directions. ," *Lecture Notes in Computer Science Vol. 1329*, pp. 363-372., 1997.
- [53] R.G., MARSTON, J.R., LOOMIS, J.M., KLATZKY, R. L. GOLLEDGE, "Stated preferences for components of a personal guidance system for nonvisual navigation ," *Journal of Visual Impairment and Blindness*, pp. 98, 135-147., 2004.
- [54] E., RIESER, J., HILL, M., HALPIN, J., and HALPIN, R. HILL, "How persons with visual impairments explore novel spaces: Strategies for good and poor performers. ," *Journal of Visual Impairment and Blindness.* , pp. 295-301. , 1993.
- [55] R.A. ZWAAN, "Situation models in language and memory. ," *Psychological Bulletin* , pp. 123. 162-185. , 1998.
- [56] E. 1948. TOLMAN, "Cognitive Maps in rats and men. ," *Psychological Review*, pp. 55. 261-292.
- [58] G.L. ALLEN, "Principles and practices for communicating route knowledge. ," *Applied Cognitive Psychology*, pp. 14, 333-359. , 2000.
- [57] S. C. & HEIDORN, P. B. HIRTLE, "The structure of cognitive maps: Representations and processes. ," *Behavior and environment: Psychological and geographical approaches.* , pp. 170-192. , 1993.
- [59] N.A, & LEGGE, G.E. GIUDICE, "Blind Navigation and the role of technology. ," in *Engineering handbook of smart technology for aging, disability, and independence.*: John Wiley & Sons. , 2008., pp. 479-500.
- [60] N.A., BAKDASH, J.Z., LEGGE. G.E. GIUDICE, "Wayfinding with words: spatial learning and navigation using dynamically updated verbal descriptions. ," *Psychological Research* , pp. 71, 347-358., 2007.
- [61] M.A., UNGAR, S., OCHAITA, E., BLADES, M., SPENCER, C. ESPINOSA, "Comparing Methods for Introducing Blind and Visually Impaired People to Unfamiliar Urban Environments. ," *Journal of Environmental Psychology* , pp.

18, 277-287. , 1998.

- [62] Z., VECCHI, T., CORNOLDI, C., MAMMARELLA, I., BONINO, D., RICCIARDI, E., PIETRINI, P. CATTANEO, "Imagery and spatial processes in blindness and visual impairment. ," *Neuroscience and Biobehavioral Reviews* , pp. 32, 1346-1360. , 2008.
- [63] A.A., GORDON, E. G., ROY, R., OGALÉ, A. KALIA, "Assessment of Indoor Route-finding Technology for People Who Are Visually Impaired. ," *Journal of Visual Impairment and Blindness* , 2010.
- [64] A., SCHAFER, J., GANDHI, S., PULEO, E., WILSON, C., ROBERTSON, M. GANZ, "PERCEPT Indoor Navigation System for the Blind and Visually Impaired: Architecture and Experimentation. ," *International Journal of Telemedicine and Applications*, 2012.
- [65] H.B., TIAN, Q., LI, B. MAGULURI, "Detecting text in Floor Maps using Histogram of Oriented Gradients.," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. , 2013.
- [66] A.A., Legge, G.E., Roy, R., & Ogale, A. Kalia, "Assessment of Indoor Route finding Technology for People who are Visually Impaired.," *Journal of Visual Impairment and Blindness*, 2010.
- [68] M., PAZZAGLIA, F., COMOLDI, C. & BERTOLO, L. DENNIS, "Spatial discourse and navigation: An analysis of route directions in the city of Venice. ," *Applied Cognitive Psychology* , pp. 13, 145-174., 1999.
- [67] N.A. GIUDICE, "Navigating novel environments: A comparison of verbal and visual learning. ," University of Minnesota., 2004.
- [69] R. M., Jacobson, R. D Kitchin, "Techniques to collect and analyze the cognitive map knowledge of persons with visual impairment or blindness: Issues of validity.," *Journal of Visual Impairment & Blindness*, 1997.
- [80] J., Alden, A., H.P.G., Schenck, M. Brabyn, "GPS performance for blind navigation in urban pedestrian settings," in *Vision 2002*.
- [70] [Online]. <https://github.com/Estimote/iOS-SDK>
- [71] Y., Harper F.M., Dreanner S., Terveen. L., Kiesler, S., Riedl, J., Kraut, E. R Ren, "Increasing Attachment to Online Communities: Evidence Based Design," *MIS Quarterly*.
- [72] A., Zang, X., Li, B. Paladugu, "Towards Designing a Crowdsourcing Network for

- Visually Impaired Users," in *IUI* , 2010.
- [73] G Brajnik, "Beyond conformance: the role of accessibility evaluation methods," WISE, 2008.
- [74] B., Wiener, W., Welsh, R Blasch, *Foundations of Orientation and Mobility* , 2nd Ed., Ed. New York, NY: AFB Press, 1997.
- [75] L. Apostolopoulos, K. Bekris, E. Folmer N. Fallah, "The user as a sensor: Navigating users with visual impariments in indoor spaces using tactile landmarks.," in *CHI*, 2012.
- [76] K., Banovic, N., Truong, K. N. SpaceSence Yatani, "SpaceSence: Representing Geographical Information to Visually Impaired People Using Spatial Tactile Feedback," in *CHI*, 2012.
- [78] M., Nationale, E., Ima, D., & Cedex, F. P. Roux, "Automatic Registration of SPOT images and digitzed maps," , 1996.
- [77] R., Park, S., Mishra, S. R., Hong, Z., Newsom, C., Joo, H., Hofer, E., Newman, M. W. Yang, "Supporting spatial awareness and inpedent wayfinding for pedestrians with visual impairments," in *ASSETS* , 2011.
- [79] Jianghua Zheng , Błażej Ciepluch , Peter Mooney , Adam C. Winstanley, Ricky Jacob, "Campus Guidance System Based on OpenStreetMap," in *Proceedings of the 9th International Symposium on Web and Wireless Geographical Information Systems*, vol. December 07-08, Maynooth , 2009.