Christian Graf   Nicholas A. Giudice   Falko Schmid   (Eds.)

Proceedings

SKALID 2012 –

Spatial Knowledge Acquisition with Limited Information Displays

Workshop co-located with Spatial Cognition 2012
Kloster Seeon, Germany, August 31, 2012

Preface

The International Workshop on Spatial Knowledge Acquisition with Limited Information Displays (SKALID 2012) was held August 31, 2012 in Kloster Seeon, Germany, in conjunction with the biennial interdisciplinary Spatial Cognition 2012 conference. The goal of the SKALID workshop was to bring people together across a broad range of disciplines to discuss methodological, technological, and theoretical concepts, challenges, and techniques related to the design of limited information displays for use in spatial contexts. Limited information displays were broadly characterized as any interface which is restricted in its size or resolution and may encompass auditory, haptic, linguistic, visual, or multimodal information displays. Of particular interest of this workshop was to solicit perspectives that cut across multiple research domains or to leverage established theories and methods from one field in order to discuss how these approaches could provide new insights or design guidance for other disciplines.

All SKALID submissions were refereed by 3 members of the workshop's international Program Committee. This team helped ensure that all submissions were relevant to the workshop, had significant intellectual and scientific merit, and had clear and coherent exposition of material. Seven papers were accepted for presentation at the workshop and publication in these proceedings. Professor Stephen Hirtle from The University of Pittsburg was the keynote speaker. In addition to summarizing their research, each presenter was asked to pose provocative or challenging questions about their work or the field more broadly. This allowed for significant time for interactive and fruitful discussion by all workshop participants.

As is obvious from the submissions, we achieved our interdisciplinary goal. Accepted papers encompassed researchers from a broad range of disciplines, including: Computer Science, Spatial Informatics, Information Systems, Human-Computer Interaction, Psychology, and others. Topics addressed in these papers covered a broad range of basic theories, empirical evidence, and interface/hardware design and evaluation, but all were linked by an interest in limited information displays. A range of visual, non-visual, and multimodal displays were discussed, with the intended users including both sighted and blind persons. An important theme evident in many of the papers dealt with what and how spatial information should be best displayed to meet the needs and tasks of this diverse user base. The workshop topics varied from selection of environmental variables, including haptic, linguistic, and visual information sources, to development of route directions, scene descriptions, and maps which are both usable and cognitively plausible. The use of cameras, augmented reality, and crowd sourcing techniques to generate and annotate limited information displays on mobile devices was the topic of several papers. Others dealt with similar ideas based on comparing information visualization techniques or new approaches for generating dynamic haptic and multimodal maps. Some of the papers advanced new theories or approaches, others evaluated the efficacy of specific new techniques or technologies, and still others performed usability testing and behavioral experiments in order to optimize interface design, insure that the information provided was perceptually and cognitively valid, or to gauge end-user acceptance. The scenarios and environments where these limited information displays were being evaluated, and the tasks aimed to be supported, ranged from perception and learning of small-scale scenes of rooms, to navigation and cognitive map development of multi-level indoor buildings, to learning and navigation of outdoor environments, to spatial knowledge acquisition at large geographic scales.

We sincerely thank the many people who made SKALID 2012 such a success: the Program Committee, the Spatial Cognition Organizing Committee, the paper contributors, and all the participants present at the workshop.

August 2012

Christian Graf
Nicholas Giudice
Falko Schmid

Organization

SKALID 2012 was jointly organized by the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition, at the University of Bremen and the VEMI Lab, in the Spatial Informatics program, School of Computing and Information Science, at the University of Maine.

Program Co-Chairs

Christian Graf	SFB/TR 8 Spatial Cognition, University of Bremen christian@maps4vips.info www.maps4vips.info
Nicholas A. Giudice	VEMI Lab, University of Maine Nicholas.giudice@maine.edu www.vemilab.org
Falko Schmid	SFB/TR 8 Spatial Cognition, University of Bremen schmid@informatik.uni-bremen.de www.cosy.informatik.uni-bremen.de/staff/schmid

Program Committee

Susane Boll	Media Informatics, Carl von Ossietzky Universität Oldenburg
Stephen Hirtle	School of Information Sciences, University of Pittsburgh
Christoph Hölscher	Center for Cognitive Science, Universität Freiburg
Andreas Hub	Visualization and Interactive Systems Institute (VIS), Universität Stuttgart
Alexander Klippel	Human Factors in GIScience Lab, Penn State University
Amy Lobben	Spatial and Map Cognition Research Lab, Department of Geography, University of Oregon
Dan Montello	Department of Geography, University of California
Martin Pielot	Intelligent User Interface Group, OFFIS Oldenburg
Martin Raubal	Institute of Cartography and Geoinformation, ETH Zürich
Kai Florian-Richter	Infrastructure Engineering, The University of Melbourne
Holly Taylor	Department of Psychology, Tufts University
Gerhard Weber	Chair Human-Computer Interaction, Technische Universität Dresden
Stephan Winter	Department of Infrastructure Engineering, The University of Melbourne

Editors' addresses

Christian Graf & Falko Schmid
Cognitive Systems (CoSy)
FB3 - Informatics
Universität Bremen
P.O. Box 330 440
28334 Bremen, Germany

Nicholas Giudice
Virtual Environment and Multimodal Interaction (VEMI) Lab
School of Computing and Information Science, Room 348 Boardman Hall
University of Maine, Orono ME, 04469-5711, USA

Copyright Note

Table of Contents

Indoor Scene Knowledge Acquisition using a Natural Language Interface

Saranya Kesavan and Nicholas A. Giudice

School of Computing and Information Science

Spatial Informatics Program, University of Maine

nicholas.giudice@smaine.edu

Abstract. This paper proposes an interface that uses automatically-generated Natural Language (NL) descriptions to describe indoor scenes based on photos taken of that scene from smartphones or other portable camera-equipped mobile devices. The goal is to develop a non-visual interface based on spatio-linguistic descriptions which could assist blind people in knowing the contents of an indoor scene (e.g., room structure, furniture, landmarks, etc.) and supporting efficient navigation of this space based on these descriptions. In this paper, we concentrate on understanding the most salient content of a stereotypic indoor scene that is described by an observer, categorizing the description strategies employed in this process, and evaluating the best presentation of directional information using NL descriptions in order to support the most accurate spatial behaviors and mental representations of these scenes by means of human behavioral experiments. This knowledge will then be used to develop a domain specific indoor scene ontology, which in turn will be used to generate automated NL descriptions of indoor scenes based on their photographs, which will finally be integrated into a real-time non-visual scene description system.

Keywords: Natural Language, indoor scene description, indoor spatial knowledge,indoor scene ontology.

1 Introduction

Navigation involves a process of controlling and monitoring the movement of any physical entity from one place to another [1]. Humans carry out this navigation process in both outdoor and indoor environments, often with the aid of external navigation aids, such as maps or GPS-based guidance systems. While humans spend approximately 87% of their time indoors, comprising both familiar and unfamiliar indoor environments [2], real-time guidance systems only work outdoors due to attenuation of the GPS signal inside and a lack of standards for building information models [3]. Compared to outdoor travel, the lack of global landmarks and complexity of indoor environments makes the task of navigation within buildings more challenging, even with the advent of indoor navigation assistance [4]. The systems for navigation assistance that do exist are almost exclusively based on visual interfaces and thus are inaccessible to blind and low vision people, one of the fastest growing demographics of our aging population [5]. To address this information gap, this paper discusses the development of an indoor navigation and scene description system which provides non-visual access to indoor environmental information by means of Natural Language descriptions delivered with the help of smartphones.

2 Background

The majority of the extant literature and technology development on accessible navigation devices relates to technology for detecting and avoiding obstacles to the path of travel or speech-enabled GPS systems for street navigation (see [6] for review). Several systems providing non-visual access to indoor environments have also been developed [see 7 for discussion]. As with the outdoor systems based on street networks, these technologies only provide network information about corridor connectivity or give landmark descriptions [8]. Beyond providing a label for a specific location (e.g., auditorium), there is no existing technology that describes the layout or salient features of these locations (i.e., the nodes which are linked by the network). As such, a blind person may have navigation assistance when traveling a route but once they reach their destination, access to functionally useful spatial knowledge regarding this location is generally limited or non-existent (e.g., the bounding contour of a meeting room, the position and orientation of a couch in the waiting room of a doctor's office, etc.).

Where navigation assistance of the network structure of large-scale environments benefits both blind and sighted users alike, sighted individuals rarely need assistance gaining information about these small-scale environments as the information is directly perceived through visual access to the scene. However, for a blind navigator relying on non-visual sensing, which is generally more proximal and less spatially precise, we argue that lack of access to spatial information about these local environments can be equally detrimental to accurate navigation, spatial learning, and cognitive map development. To date, limited research has been conducted to investigate the description of indoor scenes or how knowledge of the spatial distribution of architectural elements and salient objects in these spaces can be best imparted to blind people through non-visual channels. The limited research that has been done in this domain has required the use of expensive wearable specialty devices for acquiring spatial information. For example [9], discusses the use of a wearable device which converts visual information into tactile signals but carrying a specialized device for this purpose requires cumbersome, expensive hardware and the use of potentially confusing sensory translation algorithms. By contrast, our goal is to develop a system based on commercially-available hardware and an intuitive, easy to understand user interface. To this end, this paper proposes a work-in-progress system that provides non-visual access to specific indoor locations (scenes) through the use of NL descriptions delivered via a smartphone.

3 Natural Language – A Limited Information Display

Natural Language represents an intuitive interface as it is innate to most humans and is easy to generate using text-to-speech engines. It is often used in spatial contexts, e.g. direction giving, and has the advantage of being equally accessible to both sighted and blind users. Owing to the sparse information content that can be specified using a serial, temporally extended, and low-bandwidth medium, NL is considered a limited information display. In addition, NL involves more cognitive load in working memory than perceptual interfaces as it requires cognitive mediation to interpret the verbal information being described, such as metric, topological, and other spatial information [10]. This paper proposes a way to effectively use this limited information medium in an accessible scene description system for blind users.

4 Behavioral Experiments

Because of recent technological advancements and promising results in the field of Natural Language processing and generation, we argue that NL is one of the most important modes of information access for incorporation in non-visual interfaces and intelligent devices used by blind people. One problem is that NL description generation primarily concentrates on the semantic and syntactic aspects of the linguistic description in order to mimic human speech patterns. However, there is little formal research on understanding the ways in which a human summarizes the information they directly perceive, especially when looking at an indoor scene, through spatial verbal descriptions in order to convey survey knowledge of the scene.

Hence, in order to generate a NL description of an indoor scene, we argue that it is important to first understand the ways in which humans would naturally describe (e.g., verbally narrate) the space. NL generated without understanding of this narration is only a formal arrangement of words into sentences which abides the syntactic and semantic rules of a language following a specific architecture. To gain this knowledge, behavioral experiments must be conducted in order to understand the logic behind the human-generated NL description of an indoor scene. These results can then be compared to descriptions generated by a machine-generated NL description to assess where differences and similarities arise. The following human experiments are proposed to address this question.

4.1 Direct Observation versus Photographic Observation of an indoor scene

The end goal is for blind persons to use photos taken with their smartphones in order to obtain information about the spatial configuration of indoor scenes, including the location of its constituent objects, delivered via NL descriptions. Photos taken using smartphone cameras will inherently have a limited field of view (FOV). Hence it is important to compare the spatial information obtained from photographic observations of an indoor scene against the spatial information obtained from direct observations of the same scene to evaluate whether this limited FOV leads to exclusion of important environmental details in the ensuing spatial verbal descriptions. A behavioral experiment was conducted to evaluate whether there is a significant difference of observation by comparing the accuracy of scene re-creation based on previously generated scene descriptions from both modes. Supporting the efficacy of camera-based photos in our system, results revealed no significant differences between spatial information acquired from human or camera-based observations or re-creation accuracy based on descriptions generated from these two modes [11].

4.2 Comparing Description Strategies

Flexibility is one of the most important features of a Natural Language. One challenge is that NL descriptions of spatial information of objects in an indoor scene could be structured in different ways following different strategies. For example, a description could begin by describing the name and spatial locations of objects in one corner of the room and then follow a cyclic clockwise strategy of describing the other objects around the room. Alternatively, a description could combine objects based on their functionality, e.g. describing the spatial location of all the tables that are present in a room, then describing the chairs, etc.

Research conducted in [12] suggests that the choice of description strategies also depends on the spatial extent being described. Hence, it is important to first identify the different scene description strategies people adopt from a common perspective and then to determine which of these strategies leads to the acquisition of the most accurate spatial information while minimizing the cognitive effort required for the process. A behavioral study is currently being conducted to understand the different types of strategies that are used by humans, and the effectiveness of each for conveying accurate indoor scene descriptions. Another behavioral study will then be conducted to investigate which among those strategies helps the user to gain the most accurate spatial information supporting spatial learning and behavior in the space.

4.3 Presentation of Directional Cues

For a linguistic scene description to generate a mental map that is comparable in function with mental maps developed from visual perception, it is extremely important to have an accurate method for specifying directional cues about the spatial locations of objects within the scene. As with scene description strategies, there are different ways to verbally present these directional cues to the user. The work done in [13] suggests that people best understand directional cues when they are presented using relative directions rather than using only absolute directions. However, we are unaware of formal research investigating the best way to present directional cues with the highest precision within a relative reference frame.

Degree measurements and clock face directions are the most common ways to present angular information using a relative frame of reference. For example, “a desk is at your 1 O’ clock position” and “a desk is at 30 degrees on your right” both specify the same spatial location of the desk. But it is important to know which of these presentation methods leads to the most accurate perception of directional information. To address this question, a behavioral study is currently being conducted to compare the accuracy of angular perception based on these two types of directional cues.

5 Machine Generated vs. Human Generated Natural Language

Natural Language descriptions of indoor scenes will be generated based on the results of the above mentioned behavioral experiments. They are expected to provide access to accurate spatial information for use by a blind user when made available. However, it is also important to compare the NL descriptions created by an automated machine with the NL descriptions created by a human user in order that results from the latter can guide development of the former. This could be tested by asking participants to reproduce the scenes based on the NL descriptions from both human generated and machine generated descriptions. The accuracy of re-created scenes should be tested for any significant differences in the ensuing re-creations before being implemented in a real time indoor scene description system.

6 Indoor Scene Ontology

Natural Language Generation architecture involves a procedural and formal way of arranging raw spatial information that must then be converted to a NL [14]. To support this process, it is important to represent spatial information of indoor scenes in a formal setting, e.g. as an indoor scene ontology. Although we have ontologies available for characterizing indoor spaces in terms of corridors and pathways [15], there are currently no ontologies available to represent indoor scenes. We argue for the importance of constructing an indoor scene ontology which represents human described scene information. The primary goal of this ontology is to formally reflect and represent the ways in which humans perceive space (from the above mentioned behavioral experiments) and to structure the relevant information into a robust and flexible NL description. For example, the envisaged indoor scene ontology should involve a saliency rating of the objects that are typically present in an indoor scene. It should also be related with the existing linguistic ontology of space as proposed in [16] in order to fill the gap between human perception and formal linguistic procedures used in NL generation. Using the NL Generation techniques mentioned in [14], a NL description of indoor scenes could be developed based on the information represented in the proposed indoor scene ontology.

7 Conclusion

This paper proposes a NL user interface for describing indoor scenes to visually impaired people. While current natural language systems concentrate on the semantic and syntactic components of natural language, we propose an automated NL system that is aimed at mimicking accurate descriptions delivered by humans in terms of spatial knowledge acquisition and information delivery as established from our human behavioral experiments. Several experiments are proposed to better understand the ways in which humans perceive and describe indoor scenes in order to establish the most salient information content and description strategies. Finally, we propose the construction of an indoor scene ontology to formally represent the knowledge acquired from the results of our behavioral experiments.

Acknowledgement: This project was supported by NSF grant CDI-1028895. Thanks also to Bill Whalen for comments on the manuscript.

References

1. Bowditch, N.: CHAPTER 1 INTRODUCTION TO MARINE NAVIGATION. The American Practical Navigator. pp. 1-14. Defense Mapping Agency Hydrographic/Topographic Center, Maryland (1995).
2. Klepeis, N.E., Nelson, W.C., Ott, et al.: The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. Journal of Exposure Analysis and Environmental Epidemiology.11, 231-52 (2001).
3. Giudice, N.A, Walton, L &Worboys, M.F.: The informatics of indoor and outdoor space: A research agenda. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness.47-53 (2010).
4. Butz, A., Kr, A., &Lohse, M.: A Hybrid Indoor Navigation System. Proceedings of the 6th international conference on Intelligent user interfaces (IUI '01). 25-32 (2001).
5. World Health Organization: WHO Visual impairment and blindness, http://www.who.int/mediacentre/factsheets/fs282/en/.
6. Giudice, N.A., &Legge, G.E.: Blind navigation and the role of technology. Handbook of Smart Technology for Aging, Disability, and Independence.(2008).
7. Kulyukin, V., Gharpure, C., Nicholson, J., Pavithran, S.: RFID in robot-assisted indoor navigation for the visually impaired. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2, 1979-1984 (2004).
8. Bentzen, B. & Mitchell, P. A.: Audible Signage as a Wayfinding Aid: Verbal Land mark versus Talking Signs. Journal of Visual Impairment & Blindness, 89, 6, 494-505 (1995).
9. Johnson, L. A, & Higgins, C.M.: A navigation aid for the blind using tactile-visual sensory substitution, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1, 6289-92 (2006).
10. Klatzky, R.L., Marston, J.R., Giudice, N. A, Golledge, R.G., Loomis, J.M.: Cognitive load of navigating without vision when guided by virtual sound versus spatial language. Journal of experimental psychology, Applied. 12, 223-32 (2006).
11. Kesavan, S., &Giudice, N.A.: Automated natural language description of indoor spaces. Proceedings of the doctoral Colloquium, COSIT '11. 1-5 (2011).
12. Tenbrink, T., & Coventry, K.: Spatial Strategies in the Description of Complex Configurations. Discourse Processes: A multi-disciplinary journal. 8, 37-41 (2011).
13. Ishikawa, T., &Kiyomoto, M.: Turn to the Left or to the West: Verbal Navigational, Proceedings of the 5th GIScience international conference, 5266, 119-132 (2008).
14. Reiter, E. & Dale,R.: Building applied natural language generation systems. Natural Language Engineering, 3, 1, 57-87 (1997).
15. Worboys, M.: Modeling indoor space. Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness - ISA '11. 1-6 (2011).
16. Bateman, J. A., Hois, J., Ross, R., &Tenbrink, T.: A linguistic ontology of space for natural language processing. Artificial Intelligence.174, 1027-1071 (2010).

Building Augmented You-are-here Maps through Collaborative Annotations for the Visually Impaired

Limin Zeng, Gerhard Weber

Technische Universität Dresden
Institut für Angewandte Informatik

{limin.zeng,gerhard.weber}@tu-dresden.de

Abstract. For the visually impaired it is important to learn about different kinds of spatial knowledge from non-visual maps, specifically while walking via mobile devices. In this article, we presented an augmented audio-haptic You-are-here map system based on a novel pin-matrix device. Through the proposed system, the users can acquire not only the basic geographic information and their location, but also augmented accessibility attributes of geographic features by user-contributed annotations. Furthermore, we at the first time discuss the annotation taxonomy on geographic accessibility, towards building a systemic methodology.

Keywords: haptic interaction, social interaction, annotation, taxonomy, you-are-here map, outdoor

1 Introduction

Apart from the geographic information, for the visually impaired they expect to know additional geographic accessibility information which is specific for them. For example, the blind prefers to know if there is a non-barrier sidewalk to the entrance of the nearby POI and where, or the types of doors (e.g., automatic or manual). However, it's time-consuming and cost-consuming to collect those kinds of accessibility geographic data over the world by one or several organizations.

Besides, due to lack of accessible location-aware YAH maps for the visually impaired, it is hard for them to explore the surrounding environments while walking outside, despite the mainstream GPS-based navigation systems would announce where users are, e.g. the name of the street or the nearby point of interest. In this paper, in addition to acquiring basic geographic information, we focus on investigating which other kinds of information would be acquired from the location-based YAH maps for the visually impaired, such as location and augmented accessibility information. We presented a tactile You-are-here map system on a portable pin-matrix device (PMD), and proposed a collaborative approach to gather accessibility information of geographic features from users' annotating. Furthermore, we discussed about users' annotation taxonomy systematically from its definition to the data model.

2 Acquisition of spatial knowledge from map exploration

2.1 Basic geographic knowledge

The basic geographic knowledge contains the spatial layout of map elements, names and categories of geographic features, and other map elements (e.g. scale, north direction). Although the swell-paper based maps have great touch perception, they only can represent a few of brief and static information, as well as related map legend in Braille. To present much more map elements and with detailed descriptions, the acoustic output has been employed in recent decades, from the auditory icons and sonification to text-to-speech (TTS) synthesis, like in [1, 2]. However, it is hard to learn about precise layout from the virtual maps. Aiming at obtaining explicit touch perception simultaneously, haptic-audio maps have been proposed on touch-screen tablet [3] and pin-matrix device [4].

2.2 Spatial relationship to users on location-aware maps

In addition to rendering basic geographic information, the location-aware maps state users' current position and the spatial relationship between them and the surrounding environments. TP3 [5] allows users to discover spatial relationship to nearby point of interests on mobile phones, e.g. distance and orientation. Specifically, the novel spatial tactile feedback in SpaceSence [6] represents explicitly the orientation from users' location to the destination. However, it is still challengeable to allow the users to explore the surroundings explicitly.

2.3 Augmented geographic accessibility

For people with special needs, the kind of geographic information stored in current map database is not enough, because they have their own additional requirements. For example, the online Wheelmap[1] and Access Together[2] collect the accessibility features of POIs in cities for wheelchair users, through user-contributed annotations. However, the visually impaired has much more special requirements than the disabled people who are sighted. In addition to avoiding various unaware obstacles, they expect to know accessibility information of geographic features while on the move. The "Look-and-Listen-Map"[3] is a project to prepare a free accessible geo-data for the visually impaired, such as traffic signals with or without sound. But due to lack of an accessible platform, the visually impaired is hard to benefit from the project currently. As one possible benefit from those collaborative accessible geo-data, it can improve the performance of involved applications, like a personalized route plan [7].

[1] www.wheelmap.org
[2] www.accesstogether.org
[3] http://www.blind.accessiblemaps.org/

3 Augmented tactile YAH maps through collaborative annotations

In this section, we described a tactile YAH maps on a mobile pin-matrix device. Through the proposed system, the users not only can explore maps and learn about their location context, but also can acquire knowledge on geographic accessibility in their cities by collaborative annotations.

3.1 The System Architecture

As shown in Figure 1, the system is a typical C/S (Client/Server) based system. Its server stores map data, annotation data and user information, and responses the map client, like sending map data. The ubiquitous mobile internet enables the client to connect the server at anywhere, and to present a tactile map with current location context. Additionally, it's convenient to read annotation on the go.

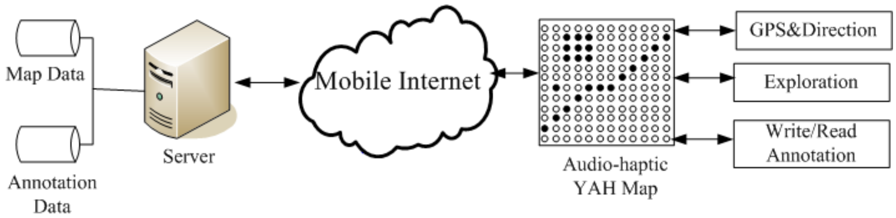


Fig. 1. The System Architecture

3.2 A Prototyped System

The Figure 2 (left) illustrates the overview of the prototype, consisting of a portable PMD, a WiiCane, a smart phone, an earphone with a microphone and a portable computer. The WiiCane is made by mounting a Wii remote controller on the top of a normal white cane. The smart phone is mounted on one shoulder, which has involved sensors like GPS, digital compass and Bluetooth. Users can listen to related annotations by the earphone. The computer in a backpack, runs the main application, and connects all of the other devices through Bluetooth or USB interface.

Besides, we design a set of tactile map symbols and YAH symbols through raised pins, to represent the YAH maps, see Figure 2 (right). In particular, the set of YAH symbols has eight symbols which point at one direction respectively, e.g. south, southwest, etc. Therefore, the YAH symbols would present users' location and heading orientation simultaneously on maps while walking or stopping.

While interacting with the YAH map, users can press the buttons on the WiiCane to panning or zooming. Due to the touch-sensitive feature of the PMD, users can obtain auditory descriptions by one finger contacting involved map symbols. To inquire "Where I AM", the YAH symbol will present in the center of the display automatically. With the help of the YAH symbols, users would explore the surroundings and discover the spatial relationship between themselves and nearby geographic features, such as the orientation and distance to a bus stop or a building.

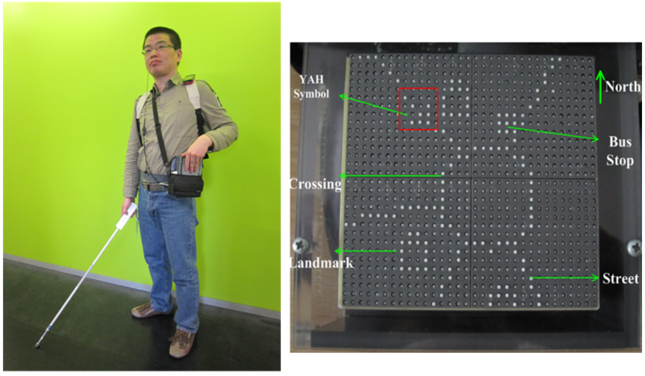


Fig. 2. The prototype of the audio-haptic location-aware YAH maps. (left: the overview of the system; right: a screenshot of YAH maps with various symbols)

4 Taxonomy of Augmented Geo-accessibility (AGa) Annotation

Although there are a couple of systems to collect enhanced accessible geo-information for the people with special needs via user-contributed annotation data, from the perspective of scientific research it is still lack of a comprehensive investigation of annotation taxonomy. Thus, in this section we focus on systematically discussing how to utilize collaborative annotated data to enhance accessibility of geographic features in real world, from its definition to data model and involved applications.

4.1 AGa Annotation Definition

"*An AGa annotation is virtual information to describe additional accessibility information of geographic features for people with special needs*."

The above definition contains several important points, which are stated as following:

At first, the focus of the annotation is about the accessibility features of geographic features in physical world. Thus, it will cover an array of items which dependent on different special user requirements and categories of existing geographic features, like auditory output of traffic signals for the visually impaired, and elevators in underground stations for the mobility impaired.

Secondly, the definition is general and simply to include a broad range of end users, who can be disabled people, but also can be temporal users like people having leg injures or wheeling a baby car.

Thirdly, the virtual information is linked to geographic referencing objects in physical world, and the descriptions of virtual information would be in a number of methods, from text, audio media, videos and pictures to haptic/tactile feedback, which can help end users to learn about involved accessibility features.

Fourthly, towards a stricter range of the term of geographic feature in the definition, it contains various existing geographic referenced objects which can be digitalized and stored in digital world, rather than all of the components on the Earth. In addition, the annotations are made not only by the end users, but also from the volunteer community who concerns about accessibility.

4.2 AGa Annotation Data Model

Different to the ubiquitous annotation systems [8] for general requirements by linking virtual information and existing objects in both physical and digital space, the dimensions of AGa annotation focus on accessibility features while reading, writing, and sharing the annotations. We list 9 dimensions to learn about AGa annotations.

1. Categories of Geographic Features: different accessibility features for different categories;
2. Location Dimension: the annotation's location in physical world;
3. Temporal Dimension: the creating/editing time;
4. Content Dimension: annotation's body is about objective mapping information or about subjective users' experiences;
5. Structure Dimension: structured accessible attributes and users' quantify annotations, e.g. rating, or an unstructured description;
6. Source Dimension: explicit annotations from users' descriptions directly and implicit annotations from digital sensors data;
7. Presentation Dimension: accessible user interfaces;
8. Editing & Sharing: annotation can be edited and shared between user groups;
9. Privacy Dimension: involved personal data;

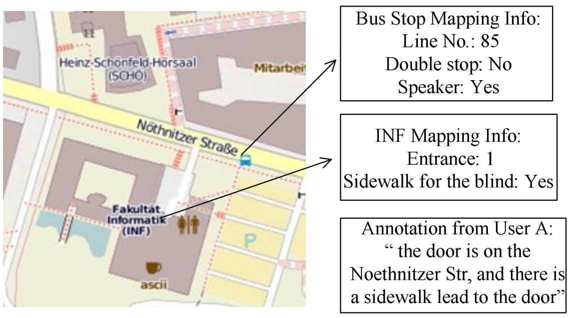


Fig. 3. An example of AGa annotations for the visually impaired

As illustrated in Figure 3, the visually impaired users would access the accessibility map information (e.g. bus stops, entrances) and annotations from others. Even if the 9 dimensions are described respectively, the relationship of them is correlative, and will impact each other in the practical applications.

5 Discussion & Conclusion

Different to rendering color-enabled maps for the sighted through the visual channel, the 2D PMD doesn't allow overlapping map symbols by the raising pins. Thus, it's important to find a suitable strategy to rendering the YAH maps on a limited portable PMD. For the visually impaired the cognitive mental maps generated while reading location-aware maps on the move might be different from a desktop-based map for pre-journey. However, what are their differences exactly is not clear yet. Besides, excepting the above mentioned 9 dimensions, which dimensions else are useful for the visually impaired?

In order to let the visually impaired acquire much more spatial knowledge on the move, the paper introduces an overview proposal how to build an augmented non-visual you-are-here maps through collaborative annotation via a portable PMD. For future work, the prototype should be evaluated with end users who are visually impaired.

References

1. Parente, P., Bishop, G.: BATs: The blind audio tactile mapping system. ACMSE. Savannah, GA, March 2003.
2. Heuten, W., Wichmann, D., Boll, S.: Interactive 3D sonification for the exploration of city maps, Proc. of NordiCHI 2006, pp. 155-164.
3. Miele, J.: Talking TMAP: Automated generation of audio-tactile maps using Smith-Kettlewell's TMAP software, British Journal of Visual Impairment, 24(2), 2006, 93-100.
4. Zeng, L., Weber., G.: Audio-Haptic browser for a geographical information system, Proc. of ICCHP 2010, pp. 466-473.
5. Yang, R., Park, S., et al.: Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments, Proc. of ASSET 2011, pp. 27-34.
6. Yatani, K., Banovic, N., Truong, K.: SpaceSense: Representing geographical information to visually impaired people using spatial tactile feedback, Proc. of CHI 2012, 415-424.
7. Völkel, T., Weber, G.: RouteCheckr: personalized multicriteria routing for mobility impaired pedestrians. Proc. of ASSETS 2008, ACM Press, pp. 185-192.
8. Hansen, F.A.: Ubiquitous annotation systems: technologies and challenges. Proc. of HYPERTEXT 2006, pp. 121-132.

Interactively Displaying Maps on a Tactile Graphics Display

Bernhard Schmitz and Thomas Ertl

Institute for Visualization and Interactive Systems, Universität Stuttgart
{Bernhard.Schmitz, Thomas.Ertl}@vis.uni-stuttgart.de

Abstract. We present a system that allows blind users to explore both indoor maps and large-scale OpenStreetMap data on a tactile graphics display. Additional information about the map object that is being touched by the user's finger is provided by text-to-speech output. Different styles allow to concentrate on specific aspects of the map. First tests show that the system can help blind users to get an impression of the layout of unknown areas, and even to get a better understanding of areas that are well-known to them.

Keywords: Tactile Maps, Tactile Graphics Display, Accessibility

1 Introduction

Being able to read maps is an important first step towards successful navigation in unknown areas. While maps are becoming ever more available for sighted persons via smartphones and services such as Google maps, blind users often still have to rely on maps that are provided specifically for them, and are often only available for small areas. At the same time, digital maps are becoming more widely available through initiatives such as OpenStreetMap. What is still missing is the link between those digital maps and the blind users. Screenreaders and Braille devices have adopted the role of this link for digital information in text format, but for spatial information such an accessibility is still missing. In this paper we present a system that is designed to provide this link, and to make maps – especially OpenStreetMap – accessible to blind users.

2 Related Work

With the advance of widespread availability of digital maps, making them accessible for all is a logical next step. Accessible maps can be classified according to several criteria, including the sensory channel that they use (e.g. tactile vs. auditory maps). Another important distinction is whether they represent a larger two-dimensional overview of the map or an approach based on a virtual observer, where only information in the immediate vicinity of a freely movable virtual observer (or cursor) is presented to the user.

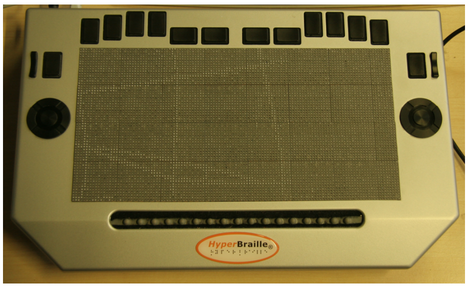


Fig. 1. The tactile graphics display, including a braille keyboard on top and two four-way digital crosses.

Purely auditory maps are normally bound to the virtual observer model, such as the auditory torch by Heuten et al. [2, 1]. In contrast to that, the works on tactile or combined tactile and auditory maps vary between those using a virtual observer model and those representing a larger overview at once. Rice et al. combine tactile and auditory feedback in their system that controls the virtual observer by a force feedback mouse [5]. Our own previous work includes a virtual observer controlled by a rumble gamepad [7].

Systems that present a larger overview often require more elaborate hardware and a setup that is tied to desktop use. Wang et al. print out a certain area of the map with a thermal embosser [8]. The printout is then placed on a touchpad, allowing audio feedback upon the touch of the user's finger. A similar approach was used by Paladugu et al. with the goal of evaluating design patterns for the production of tactile maps [4]. The system by Zeng and Weber is most similar to the one presented in this paper, also using a tactile graphics display [9]. The system uses an inbuilt GIS and renders roads as lines and buildings and other points of interest as fixed symbols from a library. Our system aims at greater flexibility by using OpenStreetMap and freely choosable styles.

3 The Tactile Graphics Display

Our system displays the maps on a Tactile Graphics Display, the "Stuttgarter Stiftplatte", with a resolution of 120x60 pins and touch-sensitive sensors for feedback about the position of the user's fingers [6] (see Figure 1). However, our system does not specifically build maps for this display, e.g. by directly activating specific pins. Instead, normal graphics output is used, and the driver converts the output for use on the Tactile Graphics Display. With this approach, the system is not limited to a specific Tactile Graphics Display, instead any that can convert on-screen graphics can be used.

Because of this approach some graphical details might be lost during the conversion. This can mostly be avoided by choosing appropriate styles for the graphics display (section 6.1).

4 Maps

Our systems displays two different kinds of maps: Highly detailed maps of buildings and small outdoor areas and OpenStreetMap data for an overview of large outdoor environments.

4.1 Detailed Maps

The detailed maps are handbuilt and were originally made for the ASBUS project, which among other goals aims at making the University of Stuttgart accessible to disabled and especially blind students by providing a navigation system. The maps show buildings with rooms and even small details like pillars and benches. The maps are stored in XML files based on the CityGML standard, but limited to two dimensions.

4.2 OpenStreetMap

For large outdoor areas where no detailed maps are available, OpenStreetMap data is used. A certain area around the current viewport is downloaded. If the user scrolls out of that viewport, new data is downloaded automatically.

OpenStreetMap is a community effort in providing maps that is a viable alternative to commercial data providers, especially regarding pedestrians: In 2011 OpenStreetMap provided a more than 30% larger street network for pedestrian navigation in Germany than the commercial TomTom Multinet 2011 database [3]. OpenStreetMap data consists of three types: Nodes, ways and relations. Nodes are simple points on the map, ways are linestrings that connect the nodes and relations can contain both nodes and ways. All three can have an arbitrary number of key-value string pairs called tags, in which the type of the object, but also any additional information is stored.

5 Interactions

As in any common map application, the user may zoom and scroll the map by using the arrow keys or the four-way digital cross of the tactile graphics display. The main additional feature is that the user can click on any object by pressing a button while having the finger on the object. Its name, function or address is then read out by a text-to-speech engine. If a handbuilt map is used, this is straightforward, as all displayed objects are named. However, in a community-based environment, in our case OpenStreetMap, the data is not always present in such a straightforward manner. Therefore, in order to be helpful to the user, the text that is read out has to be chosen more diligently. If the object is tagged with a name tag, the name is read out directly. If no name is given, a combination of the type of the object and (if available) the address is read out. The type of the object is determined by the OpenStreetMap tags in combination with strings from the styles (section 6.1), that also allow a translation into other languages. If several objects are stacked on top of each other in OpenStreetMap, the objects are read out one by one after each consecutive click, beginning with the innermost. For detailed maps, consecutive clicks are equivalent to going up one step in the GML hierarchy, so that e.g. a building's name will be announced after a room number.

Some features of the system can be accessed by a menu. After opening the menu with a keystroke, the user can navigate through the menu with the cursor keys or the digital cross. The individual menu items are read out by the text-to-speech engine. The menu allows access to the various maps (section 4), the different styles (section 6.1), and to the place search (section 6.2).

6 Additional Features

While the system as described above is functional, some additional features were implemented in close collaboration with a blind colleague, who currently uses the system most frequently. Those features can greatly enhance the usability of the system.

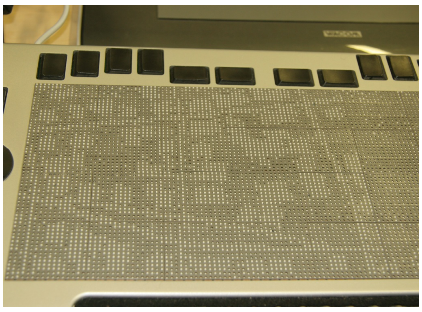
6.1 Display Styles

The Display component of our system is completely configurable. This means that for all tags in OpenStreetMap the color and thickness of the lines that are drawn, as well as the color and hatching of polygonal objects can be freely chosen. Objects can also be completely hidden, depending on their tags. Details like color and hatching cannot be reproduced on the tactile graphics display, and are mainly useful for collaboration with sighted users that use a monitor. Line thickness or the hiding of objects can be used to avoid clogging the tactile graphics display with too much information. These style settings can be stored and loaded and also added to a quick styles menu, allowing easy selection of different styles, such as "only buildings" or "only streets". Figure 2 shows a detail of an area near our University campus, showing both buildings and streets (a, b), only buildings (c, d), and only streets (e, f).

6.2 Place Search

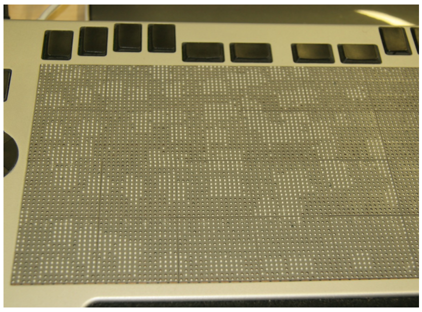
The place search allows entering the name of cities, streets or well-known entities such as landmarks. The search string is forwarded to both GeoNames and OpenStreetMap Nominatim. The places found by both services are presented to the user in an accessible list, that can again be accessed with the cursor keys. Upon selection of an entity, the map is changed to OpenStreetMap (if it is not already the case), and centered on the geographic position of the selected entity. This allows switching fast between different areas of interest.

(a) Buildings and streets (screen)

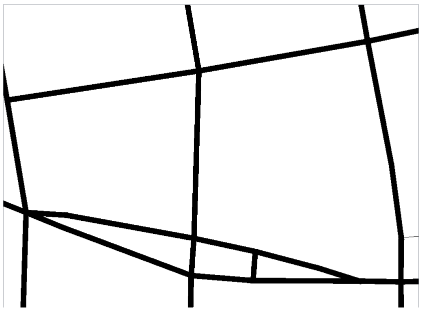


(b) Buildings and streets (tactile graphics display)

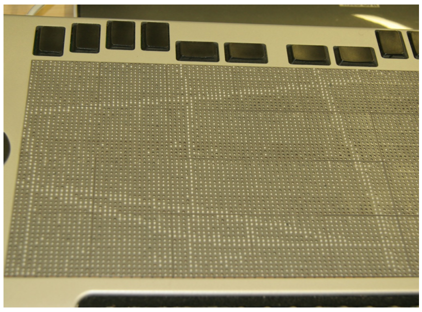
(c) Buildings only (screen)



(d) Buildings only (tactile graphics display)



(e) Streets only (screen)



(f) Streets only (tactile graphics display)

Fig. 2. A map shown with three different styles on both the computer screen and the tactile graphics display.

7 Results

First tests have shown that the system can greatly enhance the spatial understanding of its users. A blind colleague who uses the system says that she has had to correct her understanding of the street layout even of areas that she has lived in for a long time. The possibility to quickly jump to a desired location was regarded positively, as it enabled the user to go from exploring one area, such as the workplace, to another, e.g. the place of residence. Furthermore, the different styles were regarded as very helpful to reduce information overload depending on specific tasks, e.g. our colleague chose to hide all buildings when exploring the layout of streets. The use of OpenStreetMap has the advantage of providing a worldwide data set. Especially in conjunction with the flexibility achieved by the different styles, this provides a main difference from many of the previous works in this area. The effects of the low resolution can be diminished by using appropriate styles, e.g. by rendering only streets of a certain importance if zoomed out. However, the effects of using different styles for different zoom levels still need to be evaluated.

Acknowledgments. This work was funded by the Deutsche Forschungsgemeinschaft (DFG). We would like to thank Andrea Berghammer for her involvement in the design process and her invaluable feedback.

References

1. Heuten, W., Henze, N., Boll, S.: Interactive exploration of city maps with auditory torches. In: CHI '07: CHI '07 extended abstracts on human factors in computing systems. pp. 1959–1964. ACM, New York, NY, USA (2007)
2. Heuten, W., Wichmann, D., Boll, S.: Interactive 3d sonification for the exploration of city maps. In: NordiCHI '06: Proceedings of the 4th Nordic conference on human-computer interaction. pp. 155–164. ACM, New York, NY, USA (2006)
3. Neis, P., Zielstra, D., Zipf, A.: The street network evolution of crowdsourced maps: Openstreetmap in germany 20072011. Future Internet 4(1), 1–21 (2012)
4. Paladugu, D.A., Wang, Z., Li, B.: On presenting audio-tactile maps to visually impaired users for getting directions. In: Proceedings of the 28th of the international conference extended abstracts on human factors in computing systems. pp. 3955–3960. CHI EA '10, ACM, New York, NY, USA (2010)
5. Rice, M., Jacobson, R.D., Golledge, R.G., Jones, D.: Design considerations for haptic and auditory map interfaces. Cartography and Geographic Information Science (CaGIS) 32(4), 381–391 (2005)
6. Rotard, M., Taras, C., Ertl, T.: Tactile web browsing for blind people. Multimedia Tools and Applications 37(1), 53–69 (2008)
7. Schmitz, B., Ertl, T.: Making digital maps accessible using vibrations. In: Proceedings of the 12th international conference on computers helping people with special needs (ICCHP 2010). Lecture Notes in Computer Science, vol. 6179, pp. 100–107. Springer Berlin / Heidelberg (2010)
8. Wang, Z., Li, B., Hedgpeth, T., Haven, T.: Instant tactile-audio map: enabling access to digital maps for people with visual impairment. In: Proceedings of the 11th international ACM SIGACCESS conference on computers and accessibility. pp. 43–50. Assets '09, ACM, New York, NY, USA (2009)
9. Zeng, L., Weber, G.: Audio-haptic browser for a geographical information system. In: Proceedings of the 12th international conference on computers helping people with special needs (ICCHP 2010). pp. 466–473. Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg (2010)

Environmental Matching with Limited Displays

Stephen C. Hirtle and Cristina Robles

School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15260 USA
{hirtle,cmr93}@pitt.edu

Abstract. It is argued that to produce an informative spatial display on small devices one should focus on extracting distinctive features of the physical environment. These features can be communicated selectively to the user on small displays. By considering spatial, semantic, and visual information sources, one can generate cognitively adequate directions that foster spatial awareness, while limiting computational resources. This paper describes the issues involved in selecting appropriate elements within the cognitive collage of environmental spaces to generate such displays.

1 Introduction

The ability to adequately navigate using a limited information display is dependent on a variety of factors, but perhaps none as important as matching the physical environment with information in the display. It is well known that concise directions are generally preferred and easier to follow [1], but that the granularity of the directions depends on the complexity of the environment [2-3]. The problem of providing cognitively adequate directions remains a challenge for route guidance systems, which more often than not are tied to a specific level of granularity [4]. This limitation becomes more severe when the display presenting the information is limited.

One potential solution is focus on particular classes of information. It is well-established that spatial information can be viewed as a multi-level, cognitive collage in which certain kinds of information can dominate [5-8]. In this paper, we consider potential slices of the collage that can be highlighted in limited displays. We use current systems for insight into potential problem areas in applying this approach.

Thus, the focus of this paper is on environmental matching. That is, how does one take the information in the display and match it to the actual environment. For example, the simple instruction of 'turn right' at a location where five roads meet is most likely going to be inadequate and additional information would be needed to resolve the ambiguity of the instruction [9-10]. This might be through semantic information (turn right on Main St.), spatial information (make a sharp right), or visual information (turn right towards the McDonald's). Each of these alternatives is discussed in turn below.

2 Information Classes

2.1 Spatial Information

Two-dimensional spatial information is found in virtually all modern navigation systems. For example, Figure 1 from Google maps shows a path along Pocusset Street, which then turns right onto Murray Ave. That right turn that is close to 90°, as opposed to the sharp right onto Forward (westbound) or bearing slightly right onto Forward (eastbound). Having the geometry of the intersection in a heads-up display would facilitate taking the correct turn over just a simple verbal description of "turn right". However, it is interesting to note that this particular intersection, found in Pittsburgh, Pennsylvania, can also be described by the topography of the area [11]. From Pocusset Street, southbound Murray Ave is right and up the hill passing over Route 30, while westbound Forward Ave is right and down the hill passing under Route 30. This kind of the three-dimensional information is rarely provided in current navigational systems.

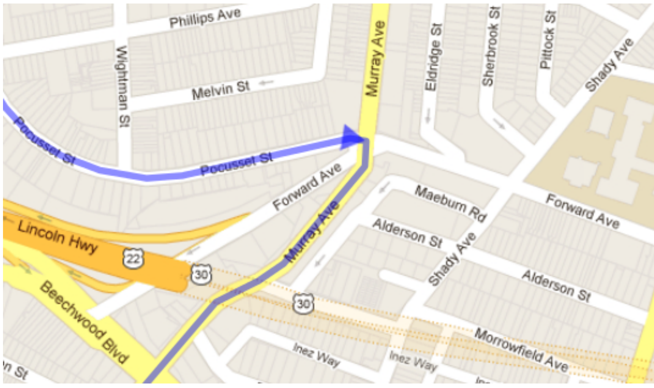


Fig. 1. Potential area of confusion for turning right in Pittsburgh, PA.

2.2 Semantic Information

Providing semantic information in terms of road names or landmarks is a second method for providing information a limited display. While typical of most current systems, problems arise when there is the signage in the environment is missing, limited, or does not match the labeling in the navigation system. For example, in Figure 1, a ramp is shown headed westbound from Forward Ave onto a shaded road labeled as "Lincoln Hwy/Route 22/Route 30", yet the physical sign onto that ramp, shown in Figure 2, reads "I-376 West Pittsburgh". The extent to which instructions match the physical environment, including what is printed on road signs, would make directions easier to follow.

Fig. 2. Visual information that does not match the labels on the map.

There is also a problem when there is limited semantic information to present to the user. Google now provides walking directions, which includes not only named streets, but also unnamed sidewalks.

Walking from the School of Information Sciences Building to Hillman Library on the University of Pittsburgh campus generates the route shown in Figure 3 with the accurate, but very confusing, set of verbal directions:

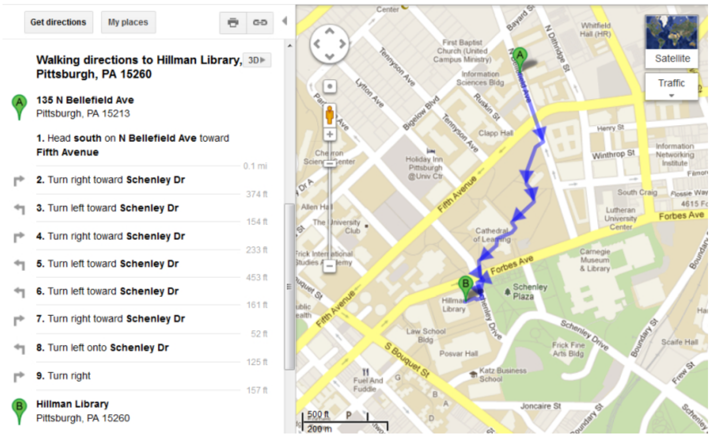


Fig. 3. Google map walking directions along a set of sidewalks

1. Head south on N Bellefield Ave toward Fifth Avenue 0.1 mi
2. Turn right toward SchenleyDr 374 ft
3. Turn left toward SchenleyDr 154 ft
4. Turn right toward SchenleyDr 233 ft
5. Turn left toward SchenleyDr 453 ft
6. Turn left toward SchenleyDr 161 ft
7. Turn right toward SchenleyDr 52 ft

8. Turn left onto Schenley Dr 125 ft
9. Turn right 157 ft [to arrive at] Hillman Library Pittsburgh, PA 15260

Pielot and Boll [12] found that such directions ignore the nature of human navigation skills and are of little use to wayfinding by pedestrians.

2.3 Visual Information

Providing visual information in terms of signage, landmarks, geographical cues, can also be helpful. In examining tricky parts of directions, Hirtle et al[4] found both cases where the visual information was obscured making directions difficult and cases where the visual information was useful in providing key information. In hand-written directions, navigators were warned, for example, when signs were obscured or where signs are missing. In contrast, landmark-rich environments were never flagged as being difficult to navigate in.

Current navigation systems rarely take advantage of visual information. One minor exception is Google Navigation for the Android, which automatically switches to street view at the end of the directions for help in locating your final destination.

3 Conclusions

It has been argued that limited information displays can successfully support spatial knowledge acquisition by providing cognitively adequate directions, which highlight the preferred knowledge of the traveler, be it spatial, visual or semantic information. This is to say, by identifying the unique attributes [13] that make up a location, an intersection or a route, one can start to build meaningful, low-resolution systems. At a first pass, such a system could be built on both user preferences and the granularity of the space [2]. In line with much of the cognitive literature, general overviews with detailed local knowledge will assist in the creation of a cognitive collage.

While the focus of this paper has been the visual display of information, it also possible to augment, or even replace, a visual display with auditory or tactile information. In these modalities it has been shown that even simple information, akin to the children's game "Hot or Cold" where feedback is given that your are on-track moving the correct direction (getting warmer) or you are off-track headed in the wrong direction (getting colder). Yang, et al [14] demonstrated that such information supports spatial awareness in pedestrians with visual impairments allowing participants to discover new points of interest and even improvise new wayfinding routes.

References

1. Daniel, M., Denis, M.: The production of route directions: Investigating conditions that favour conciseness in spatial discourse. Applied Cognitive Psychology 18, 57-75 (2004)
2. Tenbrink, T., Winter, S.P.: Variable granularity in route directions. Spatial Cognition & Computation 9, 64-93 (2009)

3. Timpf, S.: Ontologies of wayfinding: A traveler’s perspective. Networks and Spatial Economics 2, 9-33 (2002)
4. Hirtle, S.C., Richter, K.-F., Srinivas, S., Firth, R.: This is the tricky part: When directions become difficult. Journal of Spatial Information Science 1, 53-73 (2010)
5. Hirtle, S.C., Sorrows, M.E.: Designing a multi-modal tool for locating buildings on a college campus. Journal of Environmental Psychology 18, 265-276 (1998)
6. Oomes, A.H., Bojic, M., Bazen, G.: Supporting cognitive collage creation for pedestrian navigation. 8th International Conference on Engineering Psychology and Cognitive Ergonomics. LNAI vol. 5639, pp. 111-119. Springer-Verlag, San Diego (2009)
7. Sorrows, M.E., Hirtle, S.C.: The nature of landmarks for real and electronic spaces. In:Freksa, C., Mark, D. (Eds.), Spatial Information Theory, pp. 37-50. Springer, Berlin (1999)
8. Tversky, B.: Cognitive maps, cognitive collages, and spatial mental models. In:Frank, A.U., Campari, I. (Eds.), Spatial information theory: Theoretical basis for GIS (pp. 14-24). Springer-Verlag, Berlin (1993)
9. Hirtle, S.C.: The use of maps, images and "gestures" for navigation. In:Freksa, C., Brauer, W., Habel, C.,Wender, K.F., (Eds.). Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications, pp. 31-40. Springer-Verlag, Berlin (2000)
10. Klippel, A.:WayfindingChoremes. In: W. Kuhn, M. F. Worboys S. Timpf (Eds.), Spatial Information Theory: Foundations of Geographic Information Science. Conference on Spatial Information Theory (COSIT), pp. 320-334. Springer, Berlin (2003)
11. Hirtle, S.C.: Geographical design: Spatial cognition and geographical information science. Morgan & Claypool, San Rafael, CA (2011)
12. Pielot, M., Boll, S.: "In fifty metres turn left": Why turn-by-turn instructions fail pedestrians. Paper presented at Using Audio and Haptics for Delivering Spatial Information via Mobile Devices (in conjunction with MobileHCI ’10) (2010)
13. Nothegger, C., Winter, S., Raubal. M.:Selection of salient features for route directions. Spatial Cognition and Computation 4, 113-136 (2004)
14. Yang, R., Park, S., Mishra, S. R., Hong, Z., Newsom, C., Joo, H., Hofer, E., Newman, M. W.: Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments. In: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '11) pp. 27-34. ACM, New York (2011)

Verbally Assisted Virtual-Environment Tactile Maps: A Prototype System

Kris Lohmann, Matthias Kerzel, and Christopher Habel

Universität Hamburg, Department of Informatics,
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
{lohmann,kerzel,habel}@informatik.uni-hamburg.de
http://cinacs.informatik.uni-hamburg.de

Abstract. Tactile maps are non-visual substitutes for visual maps for blind and visually impaired people. In [3, 11] we suggested to increase the effectiveness of tactile maps by developing a system that generates assisting utterances to facilitate knowledge acquisition. The utterances are inspired by a corpus of assisting utterances given to tactile map explorers by human assistants. The tactile maps are realized as virtual tactile maps presented with a haptic device. The effectiveness of such a system was tested before implementation in controlled experiments reported in [10, 7]. We discuss a prototype system that was developed according to our earlier suggestions. In a user study, we positively evaluated the behavior of the prototype.

Keywords: Multi-Modal Maps, Virtual-Environment Maps, Natural Language Generation, Accessible Maps, Audio-Tactile Maps, Verbally Assisted Maps

1 Introduction

Maps are important external representations of space. For example, humans use urban-area maps in every-day scenarios such as planning a route to the next bakery or physician in environments novel to them. The access to visual maps is limited for blind and visually impaired people, therefore, tactile maps are used as substitutes. Acquiring knowledge from maps haptically is different from doing it by vision: On the one hand, information provided by tactile maps is more sparsely than that provided by visual maps; on the other hand, this information has to be integrated over time. To reduce some difficulties in haptic comprehension of maps, providing additional information via the auditory channel is useful. In [3, 11] we suggested that approaches towards audio-tactile maps can be extended by developing a system implementing *Verbally Assisting Virtual-Environment Tactile Maps* (VAVETaM). The suggested system generates situation-dependent assisting utterances, which not only inform the explorer of the tactile map about the proper names of the map objects, but include more information that reduces drawbacks stemming from the sequentiality of tactile-map reading. The generated utterances are based on a corpus study (see [9] for a discussion of the corpus

Fig. 1. The Sensable Phantom Omni and a Virtual Tactile Map

and the set of assisting utterances). For example, the assisting utterances include information about spatial relations between map objects. The Sensable Phantom Omni device allows the map user to perceive virtual tactile maps (see Fig. 1). The device can be thought of as a reverse robotic arm that generates force depending on the position of the pen-like handle. Streets and landmarks such as buildings are marked as indentations that can be felt with the device.

In this paper, we present a prototype implementation of the VAVETaM system. The goal of the development of this prototype is to show the technical feasibility of the suggested system; that is, that the prototype works. The feedback gained from the user study presented in this paper will be used for further development.

In the remainder, we discuss relevant literature (Section 2) and we briefly introduce the prototype (Section 3). Furthermore, we present a user study with the prototype (Section 4) before concluding the paper.

2 State of the Art

Audio-tactile maps as improvements of uni-modal tactile maps have successfully been developed and tested for usability. Some approaches use physical ('printout') overlay maps on touch screens (e.g., [13, 14]). The Phantom device used in our prototype has been successfully used to present virtual tactile maps with audio-feedback [12, 1, 4]. Existing approaches link sounds or brief verbal information (mostly, names) to map objects. In a corpus study, we found that humans asked to assist map explorers include more information in their assisting utterances, which can be described as brief descriptions of the local surroundings [9]. The VAVETaM approach extends earlier systems by generating assisting utterances more similar to human performance. That non-visual knowledge acquisition by direct perception of indoor environments can be facilitated by situated verbal descriptions was shown by [2].

We have positively evaluated the multi-modal interface that is described here prior to implementation in Wizard-of-Oz-like experiments (i.e., the assisting utterances were controlled by the experimenter) with both blindfolded sighted and visually impaired people [10, 7]. Participants received assisting utterances as suggested compared to a condition in which they were only informed about the names of objects. Spatial knowledge acquisition of virtual tactile maps was facilitated by verbal assistance.

3 The Prototype System

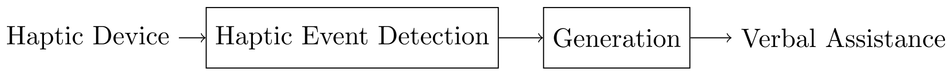


Fig. 2. From Hand Movements to Assisting Utterances

To generate assisting utterances, two important tasks have to be solved. First, the stream of position information from the haptic device is analyzed in order to detect semantic—that is, meaningful—exploration events. *Haptic event detection*, for example detects when a map user explores a street with the haptic device. This information is the input to the generation component. Second, based on this input, appropriate information is selected from a knowledge base and prepared for verbalization; that is, natural language is *generated*.

For each of these tasks, components were developed. Haptic event detection is described in more detail in [6, 5]. A generation component and the interface between event detection and generation is discussed in [9, 8].

We connected the two components to a fully working prototype system. The utterances produced by the prototype are similar to those evaluated in previous experiments (see Section 2). For example, when the movements the user performs with the device indicate that he or she is interested in a street called 'Amselweg', the system produces an output which can be translated as follows: 'This is Amselweg. This street is parallel to Blumenstraße. It forms a corner with Dorfstraße at the top end and towards the bottom it is restricted by the map frame. Furthermore, it crosses Hochstraße.' See [9], for a thorough discussion of the content of the utterances. The generation of the utterances is not based on canned text. Both, the verbalization history and the current map exploration, are taken into account: When information was already verbalized, repetitions leading to unnecessary redundancy are avoided. When the map user starts to explore another map object—for example, a landmark called 'Rathaus' (town hall)—the verbalization is stopped after the ongoing sentence is finished and the map user receives currently relevant utterances.

The generation system and the haptic event detection are implemented in Java. We use the Chai 3D toolkit for haptic rendering of the virtual tactile map. The Chai 3D toolkit and the haptic event detection are interfaced using JAVA Native Interface. Speech synthesis was realized using Mary TTS.[1]

[1] http://www.chai3d.org and http://mary.dfki.de

Table 1. System Evaluation: Translation of the Statements and Mean Response

	Translation of Statement (Original)	**Mean Response**	**SD**
A	The haptic map is understandable.	1.70	.68
B	Giving verbal assisting utterances for such maps is helpful.	1.10	.32
C	The utterances were understandable.	2.50	.97
D	I always knew exactly what is meant with the verbal utterances.	1.80	.92
E	The utterances were helpful.	1.70	.68
F	It is easy to follow the streets in the maps.	1.50	.53
G	It is easy to locate landmarks (e.g., buildings).	2.90	1.37
H	It was confusing that utterances went on when I already was at other map objects.	1.80	1.03
I	I usually knew to which map object the assisting utterances referred.	1.70	.48
J	The system behaves comprehensive.	1.50	.53

4 User Study

To evaluate the prototype system, we asked 13 participants (university students, compensated by course credit or monetary; mean age: 24.7 years, $SD = 7.2$ years, 9 males) to use the prototype and to give us feedback by indicating agreement to statements about it and by taking part in semi-structured interviews.

After an initial training with the haptic device, participants were blindfolded and interacted with the VAVETaM prototype. After they reported that they understood the behavior of the prototype, they were instructed to learn the map so that they would be able to sketch it afterwards. This instruction was included to state a clear goal for the exploration of the map. Participants could take as long as they wanted to explore and memorize the map (overall interaction time: $M = 14{:}16$ min, $SD = 4{:}46$ min; interaction after instruction: $M = 9{:}30$ min, $SD = 3{:}42$ min). After they sketched the map, they were asked to indicate agreement to a set of statements about the prototype. Selected statements are shown in Table 1.[2] They were given the list of statements in written form. The answers were given on a 1–5 Likert-type scale (1 corresponds to 'I agree completely' and 5 to 'I do not agree at all'). The list of statements was completed together with the experimenter in order to enable immediate discussions of important points (discussions were audio-recorded).

The mean responses to the statements indicate that participants considered it generally helpful to support tactile maps with natural langauge (B) and they considered the behavior of the prototype comprehensive (J). Furthermore, the virtual tactile maps were considered understandable (A, F). The responses indicate that it is possible to locate landmarks, although with more difficulty than to explore streets (G). A possible explanation is that the connected network

[2] Due to restrictions of the length of the paper, we do not discuss all statements. None of the other statements that were about the prototype produced a mean agreement worse than neutral. The full list of statements and mean responses can be retrieved from: http://www.informatik.uni-hamburg.de/WSV/VAVETaM/

of streets was easy to follow, while landmarks were not connected with each other and had to be explored one by one.

Participants considered the timing of assisting utterances appropriate to establish reference to the map object to which the assistances belonged (I). They considered the utterances helpful and they indicated that they knew what was meant (D, E). The ratings whether utterances were understandable (C) are comparably low. This seems somewhat contradictory to the high agreement with assertion D. In the discussion and the interview, participants indicated that problems with understandability were in general limited to proper names.

Currently, once sentences are started they cannot be stopped or changed anymore. From our own experience with the system, we expected that this is potentially confusing. In fact, most participants considered the inability of the prototype to change current utterances confusing (H).

In the interviews conducted subsequently, the points discussed above were elaborated and extended. Participants in general considered the set of assisting utterances appropriate and did neither miss important information, nor did they indicate that superfluous information was given. When asked for suggestions on how to improve the system, participants considered options for customization (changing the haptic presentation of the map, switching certain kinds of assisting utterances on and of) and the quality of the synthesis as potential improvements. Some participants indicated that they would have changed the content and frequency of certain utterances—however, this feedback was not consistent.

5 Conclusion

We presented the Verbally Assisting Virtual-Environment Tactile Maps (VAVETaM) prototype that solves the task to generate assisting utterances for tactile map explorations. The prototype is based on the interaction of two components: Haptic event detection and assistance generation. The effectiveness of the system was previously evaluated positively in controlled experiments [10, 7]. We presented a user study with the prototype in order to show that it works. Participants considered the prototype behavior comprehensive and the utterances and haptic presentation understandable. Potential improvements pointed out by the users are (1) the quality of the speech synthesis with respect to proper names, (2) the possibility to customize the behavior of the system, and (3) the implementation of a possibility to change ongoing utterances when the user moves to another part of the map. Aside from these potential improvements, both, with respect to the assisting utterances and to their timing, participants considered the prototype system usable, comprehensive, and helpful.

Acknowledgments. The research reported in this paper has been partially supported by DFG (German Research Foundation) in IRTG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS). We thank Martin Christof Kindsmüller for discussion. We thank the anonymous reviewers for their highly useful commentaries.

References

1. De Felice, F., Renna, F., Attolico, G., Distante, A.: A haptic/acoustic application to allow blind the access to spatial information. In: World Haptics Conference. pp. 310 – 315. Tsukuba (2007)
2. Giudice, N.A., Bakdash, J.Z., Legge, G.E.: Wayfinding with words: spatial learning and navigation using dynamically updated verbal descriptions. Psychological Research 71(3), 347–358 (2007)
3. Habel, C., Kerzel, M., Lohmann, K.: Verbal assistance in tactile-map explorations: A case for visual representations and reasoning. In: Proceedings of AAAI Workshop on Visual Representations and Reasoning 2010. Menlo Park, CA (2010)
4. Kaklanis, N., Votis, K., Moschonas, P., Tzovaras, D.: HapticRiaMaps: towards interactive exploration of web world maps for the visually impaired. In: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility. Hyderabad (2011)
5. Kerzel, M., Habel, C.: Monitoring and describing events for virtual-environment tactile-map exploration. In: Galton, A., Worboys, M.F., Duckham, M. (eds.) Proceedings of Workshop on Identifying Objects, Processes and Events. pp. 13–18. Belfast, ME (2011)
6. Kerzel, M., Habel, C.: Ereigniserkennung während der Exploration audio-taktiler Karten. In: Mensch & Computer 2012 (in press)
7. Lohmann, K.: Verbally Assisting Virtual Tactile Maps. An Interface for Visually Impaired People. Doctoral dissertation, Universität Hamburg (in preparation)
8. Lohmann, K., Eichhorn, O., Baumann, T.: Generating situated assisting utterances to facilitate tactile-map exploration: A prototype system. In: Proceedings of the NAACL Workshop on Speech and Language Processing for Assistive Technologies 2012. Montreal, QC (2012)
9. Lohmann, K., Eschenbach, C., Habel, C.: Linking spatial haptic perception to linguistic representations: assisting utterances for tactile-map explorations. In: Egenhofer, M., Giudice, N., Moratz, R., Worboys, M. (eds.) Spatial Information Theory, pp. 328–349. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2011)
10. Lohmann, K., Habel, C.: Extended verbal assistance facilitates knowledge acquisition of virtual tactile maps. In: Proceedings of Spatial Cognition 2012 (in press)
11. Lohmann, K., Kerzel, M., Habel, C.: Generating verbal assistance for tactile-map explorations. In: van der Sluis, I., Bergmann, K., van Hooijdonk, C., Theune, M. (eds.) Proceedings of the 3rd Workshop on Multimodal Output Generation 2010. Dublin (2010)
12. Moustakas, K., Nikolakis, G., Kostopoulos, K., Tzovaras, D., Strintzis, M.G.: Haptic rendering of visual data for the visually impaired. IEEE Multimedia 14(1), 62–72 (2007)
13. Parkes, D.: "NOMAD": an audio-tactile tool for the acquisition, use and management of spatially distributed information by partially sighted and blind people. In: Proceedings of the 2nd international conference on maps and graphics for visually disabled people. Nottingham (1988)
14. Wang, Z., Li, B., Hedgpeth, T., Haven, T.: Instant tactile-audio map: enabling access to digital maps for people with visual impairment. In: Proceedings of the 11th international ACM SIGACCESS conference on Computers and Accessibility. pp. 43–50. ACM, Pittsburg, PA (2009)

Using Mobile 3D Visualization Techniques to Facilitate Multi-level Cognitive Map Development of Complex Indoor Spaces

Hengshan Li and Nicholas A. Giudice

Spatial Informatics Program: School of Computing and Information Science
University of Maine, Orono, Maine 04469, USA

hengshan.li@umit.maine.edu, giudice@spatial.maine.edu

Abstract. Several studies have verified that multi-level floors are an obstacle for indoor wayfinding (e.g., navigators show greater angular error when making inter-level pointing judgments and experience more disorientation when wayfinding between floors). Previous literature has also suggested that a multi-level cognitive map could be a set of vertically super-imposed 2D cognitive maps and each level could be viewed as a region. However, little research has studied how one mentally connects / integrates the different levels of the 3D cognitive map. This paper provides new insight into how people may integrate multi-level cognitive maps based on the concept of a "transition point", a term used to represent the abstract point that connects different levels of the building. Based on transition points, we proposed the concept of simulated global indoor landmarks which are displayed on mobile devices. We predict that users can develop multi-level cognitive maps more efficiently when assisted by these global indoor landmarks. An ongoing behavioral experiment is briefly described aimed at providing empirical verification for these predictions.

Keywords: multi-level cognitive map, indoor navigation assistance, vertical information visualization, mobile information displays.

1 Introduction

From the first multi-level building of the Roman Empire to the world's highest (162-story) building, most public indoor spaces have been built based on increasingly complex indoor environments incorporating many underground levels and above ground floors. As a case in point, the growing size of malls makes these structures seem like an 'indoor city', meaning that they are large and cognitively complex environments with many possible destinations and heavy pedestrian traffic [1]. Multi-level buildings have the advantage of more efficient use of land space (particularly where space is limited or expensive), and are cheaper to cool or heat compared to a more spread-out single level structure. However, these complex multi-level buildings often cause navigators to become frustrated, disoriented, or lost during navigation (especially when traversing between floors). For instance, navigators have been shown to be significantly less accurate when pointing to locations between floors than within a single floor and inter-floor knowledge has been argued as the cause of disorientation in both physical and virtual environments [2–7]. Soeda et al. [5] demonstrated that Indoor wayfinding performance involving floor level changes is greatly hindered by disorientation during vertical travel. Likewise, Hölscher, et al. [2] reported wayfinding difficulties observed in a complex multi-level conference center, identifying incongruent floor layouts, disorienting staircases, and lack of visual access to important level-related building features as the main causes of this difficulty. Given the aforementioned literature highlighting the challenges of inter-level navigation and other relevant tasks, there is a surprising dearth of research into the underlying theory of why integrating multi-level building information is so challenging for human spatial cognition (question 1) , which is the core motivation of this paper.

Well-developed multi-level maps (whether cognitive or digital) are not only useful for the obvious applications of affording efficient inter-level indoor route planning and navigation, they could also be crucial for supporting many other scenarios. For example, in an emergency situation, firefighters needing to determine the correct location to break through a ceiling to rescue people trapped in a building, or maintenance workers needing to figure out the best route for drilling a hole to install conduit between floors. In each of these situations, a device providing perceptual information to help visualize the multi-level building structure would be extremely important for facilitating users in constructing multi-level cognitive maps which support spatial behaviors requiring integration of vertical knowledge. Therefore, we believe that the best solution to this vexing problem requires a two-pronged approach combining study of both the basic theoretical research relating to question 1 and the best interface design as assessed by question 2: how to design mobile visualization interfaces that assist tasks requiring vertical navigation, accurate learning of complex buildings, and the development of multi-level cognitive maps?

This is a position paper which aims to highlight a key problem for multi-level indoor navigation that has not been extensively studied but which represents a real and pervasive challenge given how often we are required to navigate within complex buildings. First, we provide new insights into the difficulties of inter-level indoor navigation and cognitive map development by discussing the concept of a transition point that connects different levels of a building. Next, we propose our visualization approach for integrating multi-level cognitive maps based on highlighting global indoor landmarks on mobile devices. Finally, an ongoing experiment is described that provides empirical verification of these ideas and that suggests a road map for future investigation.

2 Relevant Properties of Multi-level Cognitive Maps

Previous literature has suggested that a multi-level cognitive map could be conceptualized as a set of vertically super-imposed 2D cognitive maps having the vertical segments encoded as junctions between those maps [8], with each level being viewed as a region based on variants of the fine-to-coarse theory described in [2, 6]. However, we are not aware of any formal research that has extensively studied how one mentally connects the different levels of the 3D cognitive map. In this paper, we provide insight into how people connect these super-imposed 2D cognitive maps and introduce new visualization techniques for facilitating this process during real-time navigation.

2.1 Transition Points

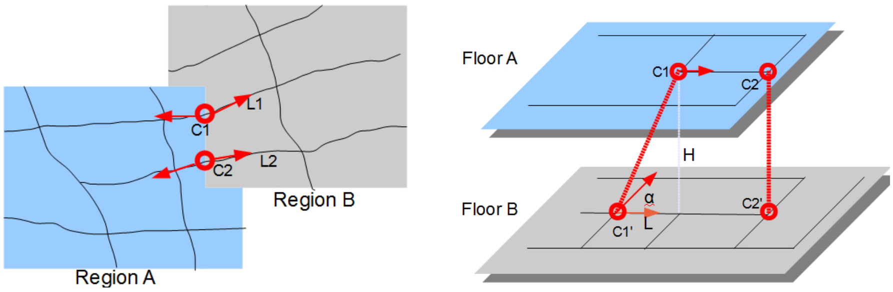


Fig. 1. Transition point in indoor and outdoor spaces

In the research by Wiener et al. [6] "region" represents perceived and encoded representations in spatial memory in which locations are grouped together and form superordinate nodes. In our research, we build on this notion by further confining the region for an indoor context as the floor's spatial extent.

The transition point represents an abstract point where navigators enter or exit a region along a route. As shown in Fig.1,an outdoor transition point is the intersection between two adjacent regions' common boundary and a route that goes through the two regions, whereas an indoor transition point is the point where users pass through a portal to enter or exit a region by an elevator or staircase. As shown in Fig.1, for outdoor space, C1 and C2 are the transition points which are the intersections between route L1 and L2 with the common boundary of region A and region B. For indoor space, there are two pairs of transition points that connect floor A and floor B (c1 with c1' and c2 with c2'). An outdoor transition point usually has two directions which are the transition point's two tangent lines' directions in two regions as shown in Fig.1, while an indoor transition point usually has one direction based on the navigator's facing direction when they get out of a portal. This notion is different from the related term, decision point, which usually refers to the intersection of corridors or travel paths [9] or the point where two route-segments meet[10], transition points are the connecting points of two regions/floors. Some indoor transition points will overlap with the decision points and therefore have several directions (e.g., one elevator has two doors or the staircase connects to a T intersection of corridors.).

When people navigate between floors, they will pass a pair of transition points (c1 with c1' or c2 with c2'). For each pair, there is a vertical transition offset H, a horizontal transition offset L, and a transition angle offset α, as illustrated in Fig.1. The transition offset H is the height between the pair of transition points located at different floors. The offset L is the distance between the transition point (e.g., C1') and the projection of the corresponding transition point (e.g., C1) on the former transition point's floor (e.g., floor B). If the two transition points are vertically aligned (e.g., an elevator connects the pair), the offset L is 0. The transition angle offset α is the difference between the directions of the two transition points. If the directions of the two transition points are the same, the offset α is 0 (e.g., an elevator connects the pair). Thus, we predicted that although multi-level indoor cognitive maps could be simplified as a 3D variant of the cognitive region, the offsets of transition points between regions/floors, particularly the horizontal and angular offsets, cause users to have greater difficulty in maintaining their spatial orientation and in developing an accurate globally coherent cognitive map of the indoor space.

2.2 Simulated Global Landmarks

The offsets between transition points illustrated above may provide insight into the known difficulties of humans in building multi-level cognitive maps based on a spatial parameter. Another potential reason for this challenge is the lack of availability of local versus global landmarks in indoor spaces [11]. Giudice et al. illustrated that the advantage of these global landmarks is that they afford an excellent fixed frame of reference which helps ground what is perceived from the local environment into a global spatial framework. However, they are often greatly reduced when learning and navigating indoor spaces. As a consequence, it is generally difficult to acquire survey type knowledge of the global spatial configuration of indoor spaces [11]. To eliminate this disadvantage and improve indoor navigational and representational efficiency, we proposed two types of simulated global indoor landmarks, transition landmarks and contiguous landmarks, which can be displayed on a mobile device. The goal is that access to these landmarks during navigation will facilitate user's ability to visualize the vertical structure of the space, which will in turn yield more accurate multi-level cognitive map development.

Transition landmarks are the highlighted information content displayed on mobile devices, composed of the transition points and directions of transition points as well as the lines connecting them, as illustrated in Fig. 2.

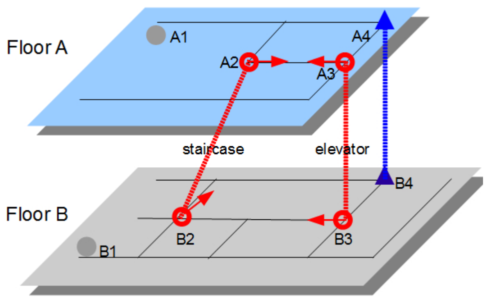


Fig. 2. Multi-level indoor global landmarks

Contiguous landmarks are also part of the highlighted information content displayed on mobile devices. They consist of vertically aligned landmarks and the lines that connect them. Landmarks are categorized as object landmarks and structural landmarks [12]. Accordingly, contiguous landmarks contain contiguous structural landmarks and contiguous object landmarks. If two floors have the same kind of structural landmarks (e.g. both floors have a cross intersection that is vertically aligned), we term it as a contiguous structural landmark. Similarly, if two floors have vertically aligned object landmarks (e.g. both of the floors have one unique blue wall at the same horizontal coordinates), we term it as a contiguous object landmark. As illustrated in Fig.2, A1 and B1 are object landmarks on each floor; A2 and B2 are the transition landmarks (staircase); A3 and B3 are both transition landmarks (elevator) and contiguous landmarks, as the transition points are located at vertically aligned T intersections; and A4 and B4 are contiguous landmarks, located at vertically aligned L intersections.

3 Experiment Design

In our research, we will experimentally evaluate whether highlighting the simulated global indoor landmarks will facilitate users' multi-level cognitive map development. In the experiment, the independent variable is the highlighting of the global landmark and there are three conditions: 1. control group: traditional 2D-based indoor maps (widely used in available indoor navigation systems); 2. birds'-eye view 3D-based indoor maps without highlighting global landmarks; 3. birds'-eye view 3D-based indoor maps highlighting global landmarks. Our hypothesis is that users in our experiment will navigate most efficiently and develop the most accurate multi-level cognitive maps with condition 3, as the global landmarks provide a fixed frame of reference in multi-level indoor spaces. Indeed, better visual access to these global landmarks is expected to facilitate improved knowledge of the landmarks interrelationship between floors and to help integrate them into a unified multi-level cognitive map.

Empirical experiments will be conducted using immersive Virtual Environments (VEs) coupled with a simulated PDA-sized screen as the visual interface to display information about navigation assistance. The advantage of using VEs is that we can leverage accurate real-time indoor positioning and tracking and easily manipulate the simulated building layouts and information content. The virtual building will be a three-level building with incongruent floor layouts and connected by confusing staircases, as the literature suggests that these factors cause the most confusion [2]. To maximize disorientation, staircases will be designed with a horizontal transition offset L and a transition angle offset α. Contiguous landmarks will be put in the environment. For example, we will make different floors have vertically aligned structural landmarks (e.g. cross -intersection).

The four main phases in the experiment are 1: *Route learning*. Participants will learn the route to each target picture with the assistance of the mobile device; 2: *Pointing criterion task*. Here we will test whether participants have successfully learned the four target locations from the first phase;3: *Unaided Navigation task*. Subjects will be asked to navigate to the picture using the shortest route; 4: *Drilling task*. Subjects will be asked to simulate "drilling" a hole to the above/lower floor or the left/right room. This task is designed to evaluate how efficiently users recall and calculate locations from the developed multi-level cognitive map.

Acknowledgments. This research was supported by the NSF under grant number: IIS-0916219.

References

1. Uzzell, D.: The myth of the indoor city. Journal of Environmental Psychology. 299–310 (1995).
2. Hölscher, C., Meilinger, T., Vrachliotis, G., Brösamle, M., Knauff, M.: Up the down staircase: Wayfinding strategies in multi-level buildings. Journal of Environmental Psychology. 26, 284–299 (2006).
3. Richardson, A.E., Montello, D.R., Hegarty, M.: Spatial knowledge acquisition from maps and from navigation in real and virtual environments. Memory & Cognition. 27, 741–750 (1999).
4. Passini, R.: Wayfinding in architecture(2nd ed.). New York: Van Nostrand Reinhold Company (1992).
5. Soeda, M., Kushiyama, N., Ohno, R.: Wayfinding in Cases with Vertical Motion. In Proceedings of MERA 97: Intern. Conference on Environment-Behavior Studies. pp. 559–564 (1997).
6. Wiener, J., Schnee, a, Mallot, H.: Use and interaction of navigation strategies in regionalized environments. Journal of Environmental Psychology. 24, 475–493 (2005).
7. Montello, D.R., Pick, H.L.: Integrating Knowledge of Vertically Aligned Large-Scale Spaces. Environment and Behavior. 25, 457–484 (1993).
8. Vidal, M., Amorim, M.-A., Berthoz, A.: Navigating in a virtual three-dimensional maze: how do egocentric and allocentric reference frames interact? Brain research. Cognitive brain research. 19, 244–58 (2004).
9. Haq, S., Zimring, C.: Just Down The Road A Piece: The Development of Topological Knowledge of Building Layouts. Environment & Behavior. 35, 132–160 (2003).
10. Richter, K., Klippel, A.: A model for context-specific route directions. Spatial Cognition IV. Reasoning, Action, Interaction. 58–78 (2005).
11. Giudice, N.A., Walton, L.A., Worboys, M.: The informatics of indoor and outdoor space: a research agenda. 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness. pp. 47–53 (2010).
12. Stankiewicz, B.J., Kalia, A. a: Acquisition of structural versus object landmark knowledge. Journal of experimental psychology. Human perception and performance. 33, 378–90 (2007).

A Quantitative Evaluation Approach for Cognitive Maps of Blind People

Mei Miao, Gerhard Weber

Technische Universität Dresden, Institut für Angewandte Informatik,
Nöthnitzer Straße 46, 01187 Dresden, Germany
{mei.miao, gerhard.weber} @tu-dresden.de

Abstract. In the field of Human-Computer Interaction, several projects aim to develop new technological aids, which enable and provide people who are blind the ability to navigate themselves around on their own. In order to evaluate these technological aids, the created cognitive maps which are built by the help of technological aids should be evaluated at first. A new approach has been developed for evaluating these reconstructed cognitive maps quantitatively. In this paper, we describe how this approach has been developed. Nine criteria are identified and weighted with help from the blind people. Using weighted Euclidean distance enables the cognitive maps to be compared with each other.

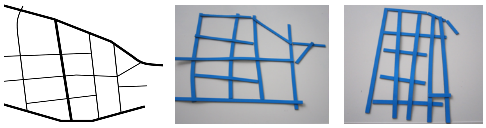
Keywords: cognitive map, blind people, quantitative evaluation.

1 Introduction

Cognitive maps also referred to as mental maps, express the essential structure of spatial information through learning processes [1]. One of a cognitive map's functions is to support navigation. Research into cognitive maps is particularly useful to urban planners, mobility specialists, and navigation aid designers [2]. For most of the blind people, it is usually impossible to travel independently. Therefore, in the field of Human Computer Interaction, several projects [3-5] aim to develop new technological aids to provide blind people the ability so that they can navigate indoor/outdoor on their own.

In order to evaluate such technological aids, the blind people are often asked to create cognitive maps which are built by the help of technological aids. Two themes will be mostly measured: route knowledge and configurational knowledge [6]. Route knowledge means the knowledge of a route from point A to point B. Configurational knowledge means the knowledge of where the roads and landmarks are located. For blind people, there are two main ways to create cognitive maps with respect to route knowledge and configurational knowledge. We can ask them to verbally describe it [7] or to reconstruct it with help of things like Modelling kit, whiteboard, bar magnets [7, 8]. Sketch mapping is the most common method for sighted people, but it is rarely used for blind people, because most of blind people are unfamiliar with it.

For evaluating the technological aids, the reconstructed cognitive maps should be evaluated. Figure 1 shows 3 maps: the original map (left, the street names are not displayed), cognitive map 1 reconstructed by subject A (middle), and cognitive map 2 reconstructed by subject B (right). How can we evaluate these maps with respect to configurational knowledge?



original map created map 1 created map 2

Fig. 1. Examples of cognitive maps created by blind people

In this paper, a new approach will be presented to quantitatively evaluate cognitive maps which are reconstructed by blind people.

2 Related Work

Until now, there has been comparatively less attention paid towards the evaluation approach of cognitive maps for the blind than for the sighted people. [9, 10] describe the methods of evaluating cognitive maps quantitatively. However, the considered cognitive maps were drawn from sighted people. In the following, we just discuss the evaluation of cognitive maps of blind people. With technological aids, blind people build the cognitive maps just based on auditory and haptic cues, so their reconstructed cognitive maps differ somewhat from those based on vision. For evaluating these cognitive maps, there is no systematic approach available. In this context, the study in [11] should be mentioned. In their study, the blind subjects were asked to model the layout of labyrinths (see fig. 2) after having explored them.

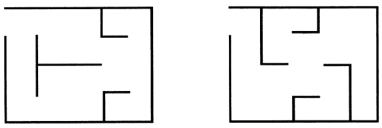


Fig. 2. Labyrinth 1 (left) and labyrinth 2 (right) [11]

The cognitive maps were analyzed in five different criteria: Number of elements: 3 in Labyrinth 1 and 4 in Labyrinth 2. Form of elements: referring to the correct identification of the L and T forms in the composition. Position: referring to the correct orientation of the elements. Placement: referring to the correct distance between the elements. Symmetry: referring to the correct axial disposition of the elements in the first labyrinth and the correct central disposition in the second labyrinth.

A map of the real environment is more complex than the labyrinths in figure 2. There are features such as street names, curves on the streets, different directions, crosses and so on. Therefore, to evaluate reconstructed cognitive maps of the real environment, we need a more complex approach.

3 Development of the Approach

In this section, we describe how this approach has been developed in detail. There are 3 steps: in the first step, the criteria were identified with which we can evaluate the cognitive maps quantitatively; then, we weighted these criteria; and finally, the method for the quantification of the criteria was chosen.

3.1 Identification of Criteria

In this step, we analysed several cognitive maps which were reconstructed by blind people in a previous study of us. In this study, ten blind people were asked to construct mental maps after learning three different maps of unknown environments using three different media (tactile map, iPad and tactile pin device (www.Hyperbraille.de)). Space restrictions do not allow to go into details here. Nine criteria were identified by the principle: one mistake should not be counted twice. The criteria were classified in 4 categories:

- Category 1: number of elements
  - *Number of correct street segments (C1)*: streets are divided in segments by crosses or branches. Only the correct street segments should be counted.
  - *Number of correctly remembered street names (C2)*: This criterion is just relevant, if we want to evaluate how easy blind people can remember the street names with the help of technological aids.
- Category 2: property of streets. Property of streets refers to the shape, name and direction of the streets.
  - *Number of correct street shapes (C3)*: referring to, if a street is straight or has curves, the number and the directions of the curves.
  - *Number of correctly assigned street names (C4)*: These include the street names which were not only remembered correctly, but also assigned to the right streets correctly.
  - *Number of correct street direction (C5)*: A tolerance range for the estimation of direction is necessary for blind people. On the basis of the 12 division on a clock face, we define 2 divisions as tolerance range. In the case that a street has curve(s), we define the direction with reference to the start and end point of the street, if there is no landmark in between.
- Category 3: arrangement of streets
  - *Number of correct crosses and branches (C6)*: a cross or branch is correct, if it is found in the original map.
- Category 4: number of errors
  - *Number of none existing streets (C7)*: streets should be counted which do not exist in the original map.
  - *Number of none existing crosses and branches (C8)*: this refers to the crosses and branches which are caused by streets which are reconstructed too long, but not by none existing streets.

- *Number of displacements of streets (C9)*: it refers to the relative position of the streets. For example, if street A crosses street B after the crossroad, it is a displacement, meaning that street A should cross street B before the crossroad.

It should be pointed out that the length belongs to the property of a street as well. However, it cannot be used as a criterion extra. Because if a street is reconstructed too long, then the length will be taken into account in the criterion "Number of none existing crosses and branches". Otherwise, if a street is reconstructed too short, then it will be taken into account in the criterion "Number of correct crosses and branches".

3.2 Weighting of the Criteria

All of the nine criteria deal with *crosses/branches* (*C1, 6, 7, 8, 9*), *street names* (*C2, 4*), *street shapes* (*C3*), or *street directions* (*C5*). In this step, we weighted these four items with weights w_1, w_2, w_3 and w_4 by involving blind people. The weights are specified as follows:

(1) At first, we had to find out if it is necessary to give the 4 items different weights. According to the relevance for getting an overview of an environment, 21 blind people were invited to rank the 4 items in order. Two different items can have the same ranking. Then, the four items are ordered by the average of the rankings from blind people: *crosses/branches* (rank at 1), *street names* (rank at 2), *street shapes* (rank at 3), and *street directions* (rank at 4).

(2) Then, we tested if the rankings of the items are significantly different. The frequency distribution of the four items and rankings was displayed in a 4x4 - contingency table. The items were paired (*crosses/branches* with *street names*, *street names* with *street shapes*, and *street shapes* with *street directions*) and tested by using the Chi-square test for homogeneity. The test showed a significant difference between *crosses/branches* and *street names*, *street shapes* and *street directions*, but not between *street names* and *street shapes* ($df = 3$, $p=0.95$, $\chi^2_{3;0.95} = 7.81$). However there is a significant difference between *street names* and *street directions*. Therefore, *street names* and *street shapes* should get the same weight.

(3) Finally, we calculated the weights w_1, w_2, w_3, and w_4. The weights are calculated according to the frequency of rank 1 and 2 of the four items. However, the frequency of rank 1 should be given more weight than the frequency of rank 2. Therefore, the frequency of rank 1 should be multiplied by a factor f ($1 < f < 3$). According to the rankings from blind people, the item *crosses/branches* should get the highest weight. If the frequency of rank 1 is multiplied by 3, the item *street names* will get the highest weight. f should be therefore less then 3. In our study, we set f equal 2 (see table 1).

Table 1. Weighted frequency distribution of the four items (a_i) of rank 1 and 2 (*H*: frequency)

rank \ **items**	a_1	a_2	a_3	a_4	Σ
1	$H_{11}*2$	$H_{12}*2$	$H_{13}*2$	$H_{14}*2$	$H_{1.}*2$
2	H_{21}	H_{22}	H_{23}	H_{24}	$H_{2.}$
Σ	$H_{11}*2+H_{21}$	$H_{12}*2+H_{22}$	$H_{13}*2+H_{23}$	$H_{14}*2+H_{24}$	$H_{1.}*2+H_{2.}$

The weights were calculated as shown in equation (1). So we got following results: $w_1 = 0{,}31$, $w_2 = w_3 = 0{,}28$, $w_4 = 0{,}13$.

$$w(a_i) = (H_{1i} * 2 + H_{2i})/(H_{1.} * 2 + H_{2.}) \quad (1)$$

3.3 Choice of Method for Quantification

For quantifying the criteria we chose the approach *weighted Euclidean distance*.

$$d(x,y) = \sqrt{\sum_{i=1}^{i} w_i * (x_i - y_i)^2} = \sqrt{w_1 * (x_1 - y_1)^2 + \Lambda + w_i * (x_i - y_i)^2} \quad (2)$$

where $x = (x_1, x_2, \ldots x_i)$, $y = (y_1, y_2, \ldots, y_i)$, w_i = weights, and $d(x, y)$ = the distance from point x to y. In our case, x_i is the *i-th* value of the original map ($x_{7, 8, 9} = 0$), while y_i is the *i-th* value of the reconstructed cognitive map, and w_i is the specified weight for *i-th* criteria (see table 2). $d(x, y)$ is the distance between the original map and cognitive map. In other words, we compare these two maps by measuring the distortion between the original map and the cognitive map. It gives a measure of how similar cognitive map to the original map actually is. The smaller the value, the more similar the cognitive map to the original map is.

Table 2. Criteria and their weights for evaluating cognitive map

criteria	weight	Original map	Cognitive map
Number of correct street segments	0.31	x_1	y_1
Number of correct remembered street names	0.28	x_2	y_2
Number of correct street shapes	0.28	x_3	y_3
Number of correct assigned street names	0.28	x_4	y_4
Number of correct street direction	0.13	x_5	y_5
Number of correct crosses and branches	0.31	x_6	y_6
Number of none existing streets	0.31	*0*	y_7
Number of none existing crosses and branches	0.31	*0*	y_8
Number of displacement of streets	0.31	*0*	y_9

As mentioned, if we do not want to evaluate how easy blind people can remember the street names with the help of technological aids, we can set its (*C2*) weight 0.

4 Conclusions

This approach has been developed within a study in which the arrangement of streets was tested with different technological aids. There were no landmarks on the original map. It is not intended for evaluating cognitive maps according to the structure of buildings such as an airport. In this case the location of landmarks should be tested, the criteria have to be extended and the weight specified. In addition, we also found out that the blind people weighted the 4 items in matters of wayfinding differently than in matters of getting an overview of an environment. This indicates that if route knowledge is tested, we need other criteria and other weights for them.

Acknowledgements

We would like to thank all of the blind participants and collaborators for spending their time on our study. We would also like to thank Limin Zeng for the preparation of the study and Michael Schmidt for useful discussions.

References:

1. Lloyd, R., Cammack, R.: Constructing cognitive maps with orientation biases. In: Portugali, J. (eds) The Construction of Cognitive Maps. pp. 187-213, Netherlands (1996)
2. Kitchin, RM.: Cognitive maps: What are they and why study them? Journal of Environmental Psychology, 14, 1-19 (1994)
3. Spindler, M., Weber, M., Prescher, D., Miao, M., Weber, G.: Translating Floor Plans into Directions. In Proceedings of the International Conference on Computers Helping People With Special Needs (ICCHP'12) (2012)
4. Daunys, G., Lauruska, V.: Sonification System of Maps for Blind – Alternative View (Vol. 2). Universal Access in Human-computer Interaction: Intelligent and ubiquitous (2009)
5. Zeng, L.; Weber, G.: Audio-Haptic Browser for a Geographical Information System. In: K. Miesenberger et al. (Eds.): Proceedings of the International Conference on Computers Helping People With Special Needs (ICCHP'10), LNCS 6180, 466-473 (2010)
6. Kitchin, R.M., Jacobson, R.D.: Techniques to Collect and Analyze the Cognitive Map Knowledge of Persons with Visual Impairment or Blindness: Issues of Validity. Journal of Visual impairments & Blindness, July-August. pp.360-376 (1997)
7. Lahav, O., Mioduser, D.: Blind Persons' Acquisition of Spatial Cognitive Mapping and Orientation Skills Supported by Virtual Environment. In Proceedings of 5th International Conference on Disability, Virtual Reality and Associated Technologies. UK (2004)
8. Ohuchi, M., Iwaya, Y., Suzuki, Y., Munekata, T.: Cognitive-Map Formation of Blind Persons in a Virtual Sound Environment. In Proceedings of the 12th International Conference on Auditory Display, London, UK (2006)
9. Waterman, S., Gordon, D.: A Quantitave-Comparative Approach to Analysis of Distortion in Mental Maps, Annals of the Association of American Geographers. pp. 326-337 (1984)
10. Peake, S., Moore, T.: Analysis of distortions in a mental map using GPS and GIS. Presented at SIRC 2004-16th Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand, November 29-30 (2004)
11. Passini, R., Proulx, G., Rainville, C. The spatio-cognitive abilities of the visually impaired population. Environment and Behavior 22, pp. 91-118 (1990)

Author Index