Enriching Navigation Tools through Human Annotations

Cole Gleason
cgleason@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA

Kris M. Kitani
kkitani@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA

Jeffery P. Bigham
jbigham@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA

ABSTRACT
People with vision impairments desire to interact independently with the physical environments in which they live, work, or travel. Accessibility research has primarily addressed this large problem-space through the creation of tools that provide route guidance to a destination. Other tasks, such as leisurely exploration or answering questions about physical environments, are not well supported by this type of tool. How can someone who is blind easily discover the historical significance of the statue in the town square? What changed since their last visit here? Well-structured data is not yet available to build applications for these use cases, and general approaches to labelling natural scenes are not robust for everyday use. Instead, accessibility researchers should generate data about physical environments by gathering descriptive and nuanced human annotations. Whether provided by volunteers, application end-users, or paid crowd workers, these annotations can enrich navigation applications to support a variety of user tasks.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '19 Workshop: Hacking Blind Navigation, May 04–09, 2019, Glasgow, UK
© 2019 Copyright held by the owner/author(s).
KEYWORDS
People with vision impairments, navigation tools, assistive technology, crowdsourcing

ACM Reference Format:

BACKGROUND AND MOTIVATION
Accessibility researchers have sought to create assistive technology that provide people with vision impairments information about the physical world and allow for independent navigation. The primary goal of many, if not most, is to support route guidance through directions from a current location to the desired destination, similar to commercial applications like Google Maps. However, navigation tools could provide much richer information about the physical environments that people inhabit or visit. For example, exploring a neighborhood with rich visual landmarks (e.g., statues, murals) is currently not well supported by applications that provide turn-by-turn directions.

Navigation applications should be expanded to support a rich variety of user tasks, including route guidance, exploration, and question-answering, as this would make many outdoor and indoor areas more accessible. In commercial and research navigation applications, route information and business points of interest are common, but other smaller and useful objects such as seating or trash cans are not typically included. Other forms of information that could be provided, such as changes to the environment or layout information is typically visual and inaccessible to blind pedestrians. While the infrastructure to locate the user outdoors is common (GPS) and it is maturing for indoor spaces (Bluetooth beacons, WiFi, etc.) [1], the sources of geographic data are limited. There is no indoor equivalent to outdoor street maps or labelled points of interest. Even outdoors, structured data is typically available for businesses or well-known attractions, but not for functional items such as seating or historical plaques. This lack of structured data about visual information in physical environments is a serious impediment to making navigation apps useful beyond route guidance.

To enable progress in the development of navigation tools, accessibility researchers should focus on methods to generate descriptions of visual information in physical spaces, whether that is a poster in the hallway or an event down the street. For important problems where the visual information is constrained, such as locating common street signs, automated approaches (i.e., object recognition) work well. However, we need richer descriptions for unique artwork or complex information, such as the layout of a room. In these cases, annotations provided by humans are likely to be more descriptive and accurate than automated approaches.

My past and current research aims to generate rich and detailed descriptions of physical objects and spaces that include everything from historical information to changes that have occurred since
the user last visited. Prior work has shown that online, remote crowdworkers can label sidewalk accessibility issues [6] or bus stop landmarks [4]. For a different domain, cycling, Torre et al. organized on-site volunteers [7] to provide detailed tags and notes for geographic areas. By combining input from both local and remote volunteers or workers, we can generate rich descriptions at scale that still include specific local knowledge. Research challenges remain in defining the scope of descriptions to be gathered, a means to share them with application developers, and methods for user interaction with these descriptions. If the community is able to address these ecological issues, future navigation tools could support new use cases to make physical spaces much more accessible to people with visual impairments.

EXAMPLES OF PAST AND CURRENT RESEARCH

Our past research has primarily focused on how we can improve the infrastructure surrounding navigation tools for people with vision impairments. My previous research, VizMap and FootNotes, explore how we can collect and display annotations in 3D space using a combination of online crowd workers and local volunteers. My current research attempts to move beyond annotations about static objects in the environment, and provide information about how an area has changed between visits.

VizMap

As seen in Hara et al. [4–6], online crowd workers can be employed to label large geographic areas for accessibility purposes, provided there is sufficient visual data of that area and an interface for labeling. As Google Street View does not provide information about building interiors, I developed VizMap to use crowdsourced videos to generate map content [3]. Videos collected by local volunteers with smartphones are processed to create a 3D reconstruction of the space using Structure from Motion. Simultaneously, online crowd workers draw bounding boxes around objects and label them in the space. These can be simple objects, such as a chair, or descriptions of unique objects like a statue. These object labels are then re-projected back into the 3D space, and the resulting labels can be used by navigation applications with their existing localization method.

FootNotes

Online crowd workers can effectively label some objects, but rich descriptive and background information is best provided by the on-site volunteers. FootNotes collects rich, textual descriptions of objects and spaces from both on-site and online volunteers, which are then embedded in OpenStreetMaps [2]. The annotations supported by FootNotes focus on visual (e.g., color, dimensions), functional (e.g., layout, floor texture), historical (e.g., why is this here?, who built it?), and social annotations (messages left by one user for another). I conducted a study with 10 visually-impaired participants, who used a navigation application prototype to experience these annotations while walking around a
downtown area. The participants generally favored visual and functional annotations, but would use historical and social ones in some contexts, including leisure.

“What’s Different Here?”

The environmental information that can be made accessible by assistive technology need not be constrained to static objects, as the physical world is often quite dynamic. Much of the information that might be labelled by a system such as VizMap or FootNotes is temporal in nature due to seasons, time of day, or just change over time. A local park might be taken over for an event, or a building might be redecorated, rendering previous labels obsolete. Many changes are primarily indicated by visual cues (e.g., signs, bright colors), and therefore people with vision impairments may be unsure about what is different since their last visit. My current research seeks to help answer the question “What’s different here?” by detecting environmental changes using a camera worn by a person with a vision impairment. When the user suspects something is different about the surrounding environment, their current video footage is matched to prior visits, then sent to a crowd worker for a description of the change. If the crowd worker deems the change is significant, their description is relayed back to the user.

RESEARCH CHALLENGES

Enriching navigation applications with additional environmental information is a broad goal, and the specific examples I have pursued in past and current research are only a small portion of possible applications. However, throughout these projects I have identified several research questions for those working on navigation tools for people with vision impairments:

1. What types of environmental information, visual or not, are important for making physical areas accessible? Which of these are we currently not including in the design of navigation tools?

2. Many navigation applications rely on existing databases of geographic information, such as OpenStreetMaps or Google Maps; how can these other types of environmental information be gathered, stored, and shared?

3. Current navigation tools typically support limited interactions for route guidance tasks. If they supported additional user tasks and a breadth of environmental information, what interaction methods will allow people with vision impairments to access it without information overload?

If the accessibility research community can design solutions and pursue research related to these questions, future commercial or research developers may have the data and environment they need to innovate on navigation applications. These applications could support people with vision impairments in every use case, thus making the places they visit much more accessible.
REFERENCES


