# Closing the Gap: Designing for the Last-Few-Meters Wayfinding Problem for People with Visual Impairments

**Manaswi Saha[1]\*, Alexander J. Fiannaca[2], Melanie Kneisel[2], Edward Cutrell[2], Meredith Ringel Morris[2]**
[1]University of Washington, Seattle; [2]Microsoft Research, Redmond
manaswi@cs.uw.edu, {alfianna, mekne, cutrell, merrie}@microsoft.com

## ABSTRACT

Despite the major role of Global Positioning Systems (GPS) as a navigation tool for people with visual impairments (VI), a crucial missing aspect of point-to-point navigation with these systems is the *last-few-meters wayfinding problem*. Due to GPS inaccuracy and inadequate map data, systems often bring a user to the vicinity of a destination but not to the exact location, causing challenges such as difficulty locating building entrances or a specific storefront from a series of stores. In this paper, we study this problem space in two parts: (1) A formative study (*N*=22) to understand challenges, current resolution techniques, and user needs; and (2) A design probe study (*N*=13) using a novel, vision-based system called *Landmark AI* to understand how technology can better address aspects of this problem. Based on these investigations, we articulate a design space for systems addressing this challenge, along with implications for future systems to support precise navigation for people with VI.

## Author Keywords

Accessibility; blindness; wayfinding; landmarks.

## CSS Concepts

• **Human-centered computing ~ Accessibility design and evaluation methods** • **Human-centered computing ~ Accessibility systems and tools**.

## INTRODUCTION

According to the World Health Organization [87], there are 285 million people with visual impairments (VI), of which 39 million are blind. For this group of people, navigation can be difficult due to challenges such as obstacles, crowds, noise, or complex layouts of physical spaces [3,16,62]. Among the many navigation tools that have been developed to cater to the needs of this community, GPS-based systems [65–68] are the most popular. This presents a challenge, in that smartphone-based GPS has a horizontal accuracy of about ±5m at best [69], with potential for much worse accuracy in areas like urban canyons [43].

\*Work completed as an intern at Microsoft Research

This means that smartphone GPS can guide a user to the vicinity of their destination, but not to the precise location (*e.g.,* to the front of a building, but not to the actual door). This gap of a few meters is acceptable for people who can rely on their vision to identify their destination but creates confusion and uncertainty for people with VI. In addition to GPS inaccuracy, location-based systems rely on map data that is often inadequate to guide the user to their intended destination due to lack of granular information. Together, these imprecisions can limit blind users' sense of independence. We call this challenge the *last-few-meters wayfinding problem* (also known as the "last 10-meters/yards" problem [22,39]).

In this paper, we investigate the problem of navigation for people with VI from the lens of the *last few meters* to a destination and use landmark recognition as the central strategy for navigation, building on the typical landmark-based navigation strategy taught to people with VI by Orientation and Mobility (O&M) specialists [40]. We conducted a formative study to understand the characteristics of this problem including the challenges faced, current resolution techniques (and where they fall short), and how an ideal system might fill the gaps left by existing navigation aids. We then developed *Landmark AI*, a computer vision system intended as a design probe for understanding how technology could be employed to help users answer three common questions surfaced by participants in our formative study: (i) "*What is around me?*", (ii) "*What does this sign read?*", and (iii) "*Is this the place I am looking for?*". Using this design probe, we conducted a user study to elicit feedback and opinions on the design of applications to address some of the last-few-meters challenges that blind pedestrians experience.

As the first work to comprehensively investigate the last-few-meters wayfinding challenge, our contributions include:

- An investigation into the problem via a formative online survey with 22 participants;
- A qualitative study of our Landmark AI design probe with 13 visually impaired users;
- A description of the design space resulting from the two studies capturing the relationship between use of landmarks and other information types with a person's mobility skills, residual vision, and situational context;
- Design implications for future camera-based AI systems targeting the last-few-meters problem.

## BACKGROUND AND RELATED WORK

Onset of VI impacts a person's ability to perform day-to-day activities, and traveling independently is a core skill that must be developed [46]. As a key part of rehabilitation training, O&M specialists teach people with VI how to safely navigate indoors and outdoors. The basic O&M techniques for navigating independently include performing systematic search and trailing [46], as well as skills based on locating and using a series of physical landmarks between locations [40]. Some commonly used landmarks to ascertain location include contrasting floor textures and coverings (*e.g.*, carpet to tiled surface or concrete sidewalk to grass pavements); using sounds (*e.g.*, refrigerator humming, birds chirping, church bells, traffic) and smells (*e.g.*, laundry room odors, kitchen smells, perfume store aromas) [70]. An O&M specialist teaches a route by breaking it down into small sections and helps the VI person identify useful landmarks along them. Using the identified set of landmarks as a checklist, the person moves from one section of the route to another by taking an *action* at each section endpoint marked by a *landmark*. For example, trailing a building line until it ends (landmark) and then taking a right turn (action). In this paper, we aim to address the last-few-meters challenge by complementing these existing mobility skills with technology that supports the discovery and use of landmarks.

### Navigation Systems for the Blind

Navigation systems for the blind is a well-studied field for both indoor and outdoor navigation. Outdoor GPS-based systems [65–68], are the most widely deployed, providing features such as announcing nearby Points of Interests (POIs) and street intersections, and providing directional guidance via turn-by-turn directions [65,67] and spatialized audio [66,68]. Although they allow a user to navigate large distances, the localization accuracy is ±5m [69], preventing users from getting very close to their target POI.

Information on locating landmarks such as doors, elevators, or stairs can play a crucial role in getting to a destination successfully. However, most systems lack such granular information. Some recent work on indoor navigation systems [4,17,20,21,49,71] have looked into providing information on such semantic features of the environment [45]. For example, NavCog3 [49] uses a BLE beacon network to provide sub-meter localization accuracy indoors and information on nearby landmarks. Such systems demonstrate the usefulness of the landmark-based navigation approach; however, they require (i) additional deployment and maintenance effort to augment the physical environment (*e.g.*, with RFID sensors [20], NFC tags [21], or Bluetooth beacons [49]), (ii) significant bootstrapping cost for setting up databases of floor maps [17] and landmarks [4,17,21,72], or (iii) require a user to carry an additional/specialized device [20]. These issues reduce the scalability and applicability of existing systems in diverse environments (*e.g.*, outdoors). The BlindWays [73] smartphone app is a system that aims to address the last-few-meters challenge without augmentation; using crowdsourced clues to assist in finding transit stops. In this paper, we investigate the full breadth of the last-few-meters wayfinding challenge and evaluate a camera-based (rather than crowdsourced) solution to find landmarks. This approach could work in tandem with outdoor or indoor navigation systems without requiring custom infrastructure.

### Camera-based Systems for the Blind

Camera-based applications serve a wide array of purposes for VI users, including simple object recognition [64,74,75], text recognition [35,74,76], and search tasks [6,7]. Object recognition systems either use computer vision [74,75], human-assistance [77–79], or a hybrid of the two [80]. Human-in-the-loop systems' latency (ranging from several seconds [80,81] to several minutes [78]) may be unacceptable for many navigation tasks; hence, our focus with Landmark AI is on automated approaches.

Coughlan *et al*. [13] used a phone camera for wayfinding by utilizing computer vision to locate and read aloud specially designed signs. Subsequent systems have looked into combining phone and wearable cameras (*e.g.*, [4,55,60]) with other sensors (*e.g.*, smartphone and motion sensors [37,47,55]), or augmenting a smartphone camera [27,30]. Using computer vision and sensor fusion techniques, these systems localize, keep track of the user's path, and provide precise corrective heading instructions. However, these systems require a time-consuming and laborious process of creating a database of likely locations, landmarks, or paths and augmenting the physical environment, making them unsuitable for exploring infrequent and unknown destinations, and unscalable for open-world exploration in natural environments. In contrast, our design probe uses only a smartphone camera without augmenting the environment and provides *in situ* feedback for both familiar and unfamiliar environments.

## FORMATIVE STUDY

We conducted an online survey on how people with VI currently navigate to understand: (i) challenges they face in the last few meters, (ii) how they resolve them, and (iii) what information would aid them in resolving these challenges.

Informed by O&M literature [40,46] and a discussion with an O&M specialist, we grouped commonly used landmarks into five categories—*structural, sound, tactile, air,* and *smell*. Some landmarks may be included in multiple categories (*e.g.,* an elevator is both a structural and a sound landmark). *Structural Landmarks* are part of the physical structure of the built environment and are usually detected either via residual vision, vision of the guide dog, or haptic resistance through a cane (*e.g.,* doors, stairways, elevators, and dropped curb edges). *Sound Landmarks* such as fountains, bell towers, and elevators generate noise. *Tactile Landmarks* have a distinct texture that is easily recognizable either through direct contact or through aids such as canes (*e.g.,* carpeted surfaces, tactile domes on curb ramps). *Air Landmarks* produce some form of heat or cool air that is felt through the skin, such as HVAC units or fans. *Smell Landmarks* have a distinct aroma (*e.g.,* perfumeries, tobacconists, bakeries).

## Method

*Participants.* We recruited 22 blind participants (9 female): 15 were between the age of 31-50, four were between 18-30, and three were above 50. Participants had varying levels of residual vision: 15 were totally blind and seven had some degree of light or color perception. 13 participants used canes as their primary mobility aid, six used guide dogs, 1 used a sighted guide, and one used other aids. Most described themselves as independent travelers (16) with varying self-confidence levels (*Mdn=4, SD=*0.9), ranging from *Not at all Confident* (1) to *Extremely Confident* (5).

*Procedure.* The survey was conducted over three weeks in August 2018. It took 30 minutes to complete and participants were compensated US$25. The survey used a recent critical incident approach [19] in which we asked participants to think of a specific recent episode in which they had experienced a last-few-meters navigation challenge. We used affinity diagramming [38] to analyze open-ended responses and identify themes. For the rest of the paper, survey participants are referred to by "S" suffixed by the participant number (*e.g.*, S1) and the counts of the responses are included in parenthesis.

## Findings

### Challenges in the Last Few Meters

Participants described challenging situations including tasks such as getting to an apartment, visiting the doctor's office, and finding specific buildings within large areas like business complexes. For instance, S19 explained "*I was dropped off at a college campus and I was unable to locate the building of my scheduled interview.*" Amongst all participants, visiting the doctor's office in a medical center was the most common scenario (6). In addition, the challenge of navigating open spaces where there is a lack of helpful environmental cues was a clear theme. Examples included indoor open spaces such as airports and malls (8), spaces with complex and varied paths like parks or universities (5), and open parking lots (5). These challenges are commonly encountered, with two-thirds of participants reporting at least some degree of familiarity with problematic destinations.

In most cases, the hardest part of traversing the last few meters was finding the intended doorway (11). Participants reported this was caused by: (i) failure of existing guidance systems such as the inaccuracy of navigation technology, the limitations of guide dogs, or missing or inaccessible signage (9); (ii) finding the right door from a cluster of doors (5); (iii) transit drop-off points being far away from the actual destination (5). S8 gave an example situation where these reasons get intermixed: "*The entrance to the building was off the parking lot rather than at the end of a sidewalk and the inside was a series of doors off a long hall*."

### Resolution Techniques

Participants responded to these challenges by using sighted assistance (17), O&M skills (11), trial and error (7), technology (2), or completely giving up (2). Though participants described sighted assistance as the most common and effective technique, it was not always useful: "*We resolved it by asking people, whoever could understand English, which was not too many at the airport at that time of the morning* (S9)." Almost everyone (21) received O&M training for navigating physical spaces (known techniques included counting steps, finding landmarks, and using sensory clues such as aromas, sounds, or tactile surfaces). Trial and error was also quite common (7) as indicated by S3: "*It's a matter of feeling around until you find the actual handle of the store.*" Participants often combined these techniques: "*I usually ask for assistance from a passing pedestrian. If no one is around, I simply try all the doors until I locate the right one. It's easier if it's a restaurant or coffee shop or any store that has a distinct aroma that I can use to pinpoint the exact location.* (S11)"

### Technological Limitations

All participants mentioned using technological solutions during navigation to access information like turn-by-turn directions (9), nearby streets, intersections and POIs (9), and human assistance (*e.g.*, Aira [79]) (1). Despite these benefits, participants reported many types of failures: "*Mainly the accuracy with it not putting me exactly at my location instead putting me a few yards away.* (S12)" The most important concern with current technology (16) was imprecision in terms of localization and granularity of information (*e.g.*, floor number of the location): "*Sometimes the GPS gets really thrown off and I end up walking in circles.*" (S3). Other issues included lack of indoor navigation (3), intermittent GPS signal (2), use of headphones blocking ambient noise (2), and battery power drain (2).

### Useful Information in Resolving the Challenges

Given the recent critical incident participants reported, we asked them to rate the categories of landmarks previously defined in terms of usefulness in that situation. Across all participants, **tactile** landmarks (*Mdn=*5, *SD=*1.1) were most preferred (11). For example, S2 "*...used the grass on the [entrances] to the apartment buildings.*" **Structural** landmarks (*Mdn=*5, *SD=*1.3) and **sound** (*Mdn=*4, *SD=*1.4) were next. **Smell** (*Mdn=*3, *SD=*1.5), and **air** (*Mdn=*3, *SD=*1.1) landmarks were least mentioned amongst all participants. S11 summarized the use of the landmark types based on the usage scenarios and their primary mobility aid, "*Because I travel with a guide dog, I mostly rely on smell and sound cues when traveling, with tactile landmarks being useful if they are under my feet, and structural landmarks being helpful if I know they are there and can give my dog the command to find the landmark such as 'find the stairs'.*"

When asked about missing information that would be useful in these situations, knowing about the layout of indoor and outdoor spaces was the most popular request (9).
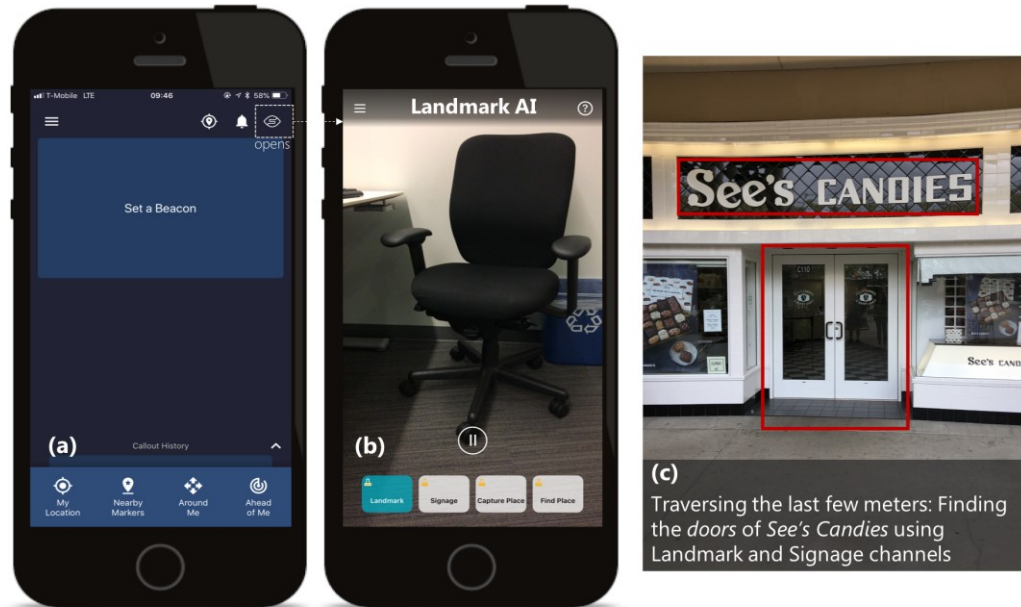
**Figure 1. Landmark AI is a vision-based app that is modeled on Seeing AI iOS app. The app is fully operable non-visually via the VoiceOver screen reader, but we show the visual UI here for clarity. (a) The Soundscape iOS navigation app helps the user get near the location. In this case, it's outside See's Candies. The top-right button is included to switch to Landmark AI once near the destination. (b) The Landmark AI app has three channels: Landmarks, Signage, and Place (a pair of Capture Place and Find Place functions). (c) Using the Landmark and Signage channels, the user can locate the entrance of See's Candies once close to the store.**

Participants also wanted to know more about existing signage (5): "*If something could identify a sign, i.e., text/logos that identify a business then that would be very helpful.*" (S6) Several participants indicated they would like to receive ego-centric layout information about nearby things (4): "*Imagine being able to walk down the hallway of an office building and hear 'men's bathroom on your left.'* (S15)." Other examples of desired information were precise auditory guidance on a smart mobile or wearable device (*e.g.,* "*approaching apartment building entrance*"), granular map information (*e.g.,* location of parking lots), and creating personal landmarks (*e.g.*, an arbitrary location like a bench in a park).

**DESIGN PROBE STUDY**
Based on our formative study's findings, we developed a vision-based app called *Landmark AI* as a design probe. We designed Landmark AI to demonstrate how landmark recognition could work, with the goal of helping participants imagine how they might use such technology combined with their existing mobility skills to overcome wayfinding problems. Because we were interested in broad questions of potential utility and context, Landmark AI was not rigorously optimized for accuracy. Rather, our probe identified categories of information for which AI developers might gather large, robust training sets such that more accurate recognition algorithms could be trained. While we do not wish to minimize the importance of accuracy and precision for such systems, these questions are out of scope for this paper. Based on this investigation, we developed a set of

design considerations for future systems addressing navigation challenges in the last few meters.

**Landmark AI System**
Landmark AI (Figure 1) is a camera-based iOS app that allows users to gather information about the space around them once they get close to a destination. It is designed to provide information that supports their existing mobility skills to aid in decision-making during navigation. We modeled our app's design on Microsoft Seeing AI [74], an iOS app that provides users with visual information via so-called *channels* (*e.g.,* reading short text, scanning bar codes, and reading currency notes). Basing our design probe on Seeing AI allowed us to minimize user training, to segregate different information types (via the channel metaphor), and to provide the user with an appropriate mental model of the feasible capabilities of current and near-term AI solutions (*i.e.*, computer vision can succeed at specific, scoped tasks such as recognizing doorways, but open-ended description of unpredictable scenes is not currently accurate). In Landmark AI we provide three new channels—Landmarks, Signage, and Places—to provide visual information that is relevant in navigating the last few meters. The app is operated by either panning the phone's back-facing camera or taking a picture (depending on the channel) to get auditory callouts.

*Landmark Channel: "What is around me?"*
Given landmarks that were indicated as useful in our formative study and prior literature on critical landmarks for navigation [45], we designed the Landmark channel to

**Figure 2. Examples of places for the Place Channel. (a) "box office counter of a theater" (b) "benches outside the ice-cream store" (c) "gummy bear outside the storefront of Margo's Sweet Shop"**

recognize structural landmarks (*e.g.*, doors, stairs, windows, elevators, and pillars) and obstacles (*e.g.,* tables, chairs, and benches) around the user as they scan the environment. Instead of choosing computationally heavy methods [10,51,54], we used a light-weight pre-trained object recognizer with reasonable accuracy ($F_1$ = 69.9 at a 99% confidence threshold for recognition) to run on commodity smartphones. The recognizer is based on the SqueezeNet [23] deep neural network model, and trained on 2,538 randomly selected images from the ImageNet database [15]. As the user pans the phone's camera, the channel announces landmarks when first recognized and every two seconds the landmark remains in the camera's view. While a real-world system would likely need to detect a much larger set of landmarks and at a much higher accuracy, constraining the detected landmarks to features common in our study location was sufficient for the purposes of a design probe demonstrating the landmark recognition concept.

*Signage Channel: "What does this sign read?"*
In our formative study, participants indicated that knowing more about nearby signage would be helpful in navigating the last few meters to a destination (*e.g.,* finding store names or signs with directions), so we designed a channel to read signage in static images the user captures with Landmark AI. An ideal implementation of this channel would perform recognition on-device in real-time [52,82], but implementing such an algorithm was out of scope for our design probe, so we used Microsoft's cloud-based Cognitive Services [83] to implement the recognition. These services require several seconds to process a frame, preventing a fully real time interaction for this channel. Despite this limitation, the signage channel gave us the opportunity to test different feedback mechanisms and study the utility of signage versus other cues when traversing the last few meters.

*Place Channel: "Is this the place I am looking for?"*
We designed the place channel to allow users to define and recognize custom landmarks. To use the channel, a user first saved a picture of a specific place they wanted to find in the future either by taking a picture themselves using *Capture Place* function or saving a picture sent from a friend (*e.g.,* a meeting place like the box office window at a theater or a specific table outside a storefront, Figure 2). The user could

then use the *Find Place* function to search for the location in the captured image. Due to the complexity of this scene matching task, we simulated this functionality in our design probe via a Wizard of Oz [36] approach, whereby a researcher would trigger customized feedback ("*<X> place found"*) when users of the design probe scanned the phone running Landmark AI over a visual scene that matched the stored scene.

**Study Method**
We conducted a three-part study using a scenario-based design [48] involving three tasks, each highlighting a last-few-meters challenge. Before using the Landmark AI design probe, users completed a set of short tutorials demonstrating the use of each channel. Each task involved getting close (~2 - 5ft) to a particular business using a popular GPS-based iOS navigation application called Microsoft Soundscape [66] and then using Landmark AI to cover the last few meters. For every task, the participants were asked to *think aloud* as they made navigation decisions. We solicited feedback on their experience including perceived utility, limitations, and design recommendations for future systems. Tasks took place within a large, outdoor two-story shopping center in the U.S. Study sessions lasted about 90 minutes, and participants also completed a demographic questionnaire. Participants were compensated US$75.

*Task 1.* *Find the elevator near the restrooms and ice-cream store*. First, participants were asked to locate the store using the GPS app and then find the elevator using their own mobility skills. Then participants were asked to walk back to the location where the GPS app stopped being useful and switch to Landmark AI to locate the elevator again. The goals of this task were to contextualize the use of Landmark AI after experiencing challenges in the last few meters when navigating on their own, and to study the use of Landmark AI in a familiar place (familiarization after completion of the first sub-task).

*Task 2*. *Find a table near the entrance of the candy shop.* In this task, participants were guided to the second floor of the shopping center and asked to use the GPS app to locate a candy shop. Participants were informed that they could switch to Landmark AI at any time. We observed how they used the two apps together, when they switched between the two, and when and why they chose to use particular channels in Landmark AI. The goal of this task was to test the usefulness of Landmark AI in visiting an unfamiliar place.

*Task 3*. *Find the box office counter for the theater.* For this task, participants were asked to imagine a scenario where they are meeting with a friend (in this case, the researcher) at the box office counter of the theatre, which the friend had sent a photo of. Their task was to locate the counter using the Place channel in Landmark AI after using the GPS app to get near the theater. The goal of this task was to understand how participants would use the more open-ended scene recognition of the Place channel.

## Participants

We recruited 13 people with VI (4 female) aged 24 - 55 (*Mean=39, SD=11*). Six participants used guide dogs, six used white canes, and one had low-vision (P3) and used magnifiers to read text. During the study, two guide dog users switched to using their canes, as they felt canes were better suited for the tasks. Participants had varying levels of functional residual vision: color perception (3), visual acuity (2), contrast sensitivity (3), peripheral vision (4), central vision (6), no vision (5), and others (2). On a 5-point Likert scale ranging from *Not at all Confident* (1) to *Extremely Confident* (5), participants had varying self-confidence levels for navigating on their own (*Mdn=4, SD=0.86*), and on a 4-point Likert scale ranging from *Not at all Familiar* (1) to *Very Familiar* (4), most participants were users of both Soundscape (*Mdn=3, SD=0.86*) and Seeing AI (*Mdn=3, SD=0.75*). Only 5 participants had some familiarity with the study location (*Mdn=1, SD=0.85*), and amongst them, none were familiar with the specific task locations.

## Data and Analysis

We audio recorded, transcribed, and coded the sessions to find general themes using deductive coding [9]. We transcribed 12 audio files; one participant's (P7) transcript was unavailable due to audio recording device failure. One researcher prepared an initial codebook based on the formative study findings and our research questions, which was later refined by a second researcher. Both researchers coded a randomly selected transcript. We used Cohen's Kappa [57] for establishing inter-rater reliability (IRR) which was 0.42 for the first iteration of the codebook, suggesting a need for more iterations [57]. We conducted three such iterations, resolving disagreements and removing or collapsing conflicting codes, before establishing IRR ($\kappa=0.69, SD=0.23$) with the final codebook. The remaining transcripts were divided and coded independently.

## FINDINGS

### Existing Wayfinding Strategies

Participants first described their wayfinding strategies included employing their O&M training, mobility aid, and residual vision (if any) to either *discover* what is around them or *search* for a specific target when they get close to their destination, depending on their familiarity with the space. Guide dogs are particularly useful in search tasks (*i.e.,* looking for a known object that the dog can recognize in the last few meters), whereas canes are more suitable for discovery tasks via O&M techniques like building trailing and structured discovery. P10 described using their residual vision to search for geometric and photometric properties of a landmark (*e.g.,* "*I can see the gleam off the metal*" or "*It looks like a big blob of colors so I think I'm in the area*") and falling back to technology when their residual vision is not sufficient: "*I usually have a monocular. […] I visually try to look around. If I get really confused, I'll go into Google Maps and zoom.*" (P10).

## Information Utility

All participants valued the information provided by Landmark AI, as the app gave access to information they might not have otherwise. They listed several reasons: ability to know what's around them, faster mobility by speeding up their search and discovery tasks, and increased independence. Participants identified several situations where they would use a system like Landmark AI: common places such as transit stops (most commonly mentioned), airports, pedestrian crossings, universities, and malls; unfamiliar places with confusing layouts such as conference centers or theaters; finding specific objects such as trash cans; and avoiding obstacles.

### Channel-Specific Utility

The Landmark channel was viewed as the most useful due to instant access to contextual information and most likely use in day-to-day life: "*I like the real time feedback. Even if it's not perfect, it's so cool because it gives you kind of a quick sense of what's around you. That's something that, as a blind person, you usually have to be pretty slow, careful, and rigorous about exploring your environment. This gives you a chance to be a little freer, or a little more spontaneous.*" (P6) Participants saw the potential to use it in different contexts by expanding the list of landmarks recognized by the system such as including restrooms and transit stops or recognizing rideshare cars' make and model. P10 described using a gummy bear statue outside the candy shop to confirm the store location; the use of residual vision with landmark detection in this case suggests that landmark detection should be designed to work in combination with users' residual vision (*e.g.,* identifying color and shape of an identified landmark) to support future visits even without the system.

The Signage channel was used to get confirmation when participants reached their destination. It was especially liked by those who had enough residual vision to detect, but not read, signs: "*I like the fact that they can pick up signage that might be too far away to see.*" (P10). The channel also provided a way to be independent, especially where Braille is unavailable. In spite of the benefits, many (5) participants found it hard to use because of difficulty in knowing *when* to look for signs (*"I didn't visually see a sign, so I didn't have a trigger to switch to sign [channel]"*—P3) and *where* to point the camera (*i.e.,* both framing the view and conceptually knowing where signs could be). To remedy this, four participants suggested detecting existence of signs in the Landmark channel and providing more guidance to capture the signs as they scan the environment.

The Place channel was the most liked channel (9) because of the ability to capture and share an uncommon location (*e.g.,* "*Meet at the bench outside Baskin Robbins in the mall*"), simplicity of usage, the wide potential of usage scenarios, and increased feeling of independence. People with residual vision found utility where their vision was inadequate: "*Because I don't have any peripheral vision, I wouldn't have noticed it [box office counter], but now that I've been there, if you said, 'Okay, go find a box office at that place.' I'd go*

*right straight to it. It's a good learning tool.*" (P10). Participants liked the channel's social focus: "*Being able to be precise and to share landmarks and to connect with people in that way, there's fun there, right*?" (P3).

### Importance of Landmarks

Amongst the landmarks currently identified by Landmark AI, the popularity of detecting doors was unanimous. Additionally, the differentiation between a door and a window was appreciated since (i) people with residual vision often have a hard time differentiating between the two due to the similarity of materials used (glass) in large shopping complexes and commercial buildings, (ii) guide dogs, who are usually good at finding doors, often get confused and lead the VI individual to a full-pane window, and (iii) cane users have trouble finding the door since they have to manually feel the door handle ("*You don't have to do that weird fumbling thing.*"—P8).

## System Design Considerations

### Seamless User Experience

Six participants liked the instantaneous feedback from the Landmark and Place channels since it gave a "*quick sense of what's around you.*" Several participants (4) felt the need for less information as it incurred cognitive load while walking. "*I really don't wanna hear everything that's coming in my way. That's too much information to process.*" (P9). They expressed the need to filter irrelevant information based on the task at hand (*e.g.*, finding the building entrance), or the situational context (*e.g.*, entering a restaurant *vs.* looking for a table) by either manually or algorithmically "*determining the importance of putting the landmark with respect to where you wanna go and what you're trying to do*"—P9.

In the design of Landmark AI, users had to switch between channels to access different types of information. Multiple participants (4) indicated the need for a simpler design favoring a merged channel to enable efficient interactions and transitions between the different types of information. Participants suggested the system should intelligently ascertain the information need based on the situational context such that the cognitive load of "*where to point and having to pick the category*" (P3) is reduced. For example, if looking at a sign, read the sign automatically instead of switching to the channel to trigger the feedback.

Physical form factor was an important consideration that was noted by several participants (4). Hands-free use was desired so as to not disrupt their navigation routine and pace. Having to hold the phone out is not ideal due to safety concerns and the difficulty of managing it along with their primary mobility aid [61]. Participants suggested using wearables such as head-mounted devices (*e.g.*, Google Glass) or wearing on-body cameras "*Because you have to hold the phone for the camera to work, I would be very limited if I wasn't using my guide dog, because she works on my left side. I can only hold the phone with my right hand. If I was using my cane, I would not be able to use this app.*" (P12). In addition to the awkwardness of using the phone, holding the phone straight was another issue. If not held straight, some participants had difficulty in understanding what was around them and where items were with respect to them; holding a tilted phone was the likely reason for their confusion.

### Accuracy

Accuracy was one of the most important concerns amongst all participants (13) as it forms the foundation for establishing trust and confidence in the system. Factors that influenced accuracy were either system-oriented or user-oriented. System oriented factors included presence of false positives in the object recognizer and lack of robustness in handling a variety of materials. For example, false positives from objects located across the glass window and false negatives caused due to environmental conditions (*e.g.*, lighting and color contrasts causing inability to read signs). User-oriented factors included difficulty in framing a well-focused image, handling users' walking pace, and perceived inaccuracy caused by not holding the phone straight and in line with their body's position. Despite the current object recognizer's inaccuracy, participants explained even with inaccurate information, they would rely on their mobility skills to support them when technology fails. An instance was confirming the information provided by the application (*e.g.*, checking a table's existence with their cane).

Closely tied to the accuracy of the object recognizer is accurately capturing the scene and receiving usable feedback. For example, participants were concerned about being too far away or too close while taking a picture. Similarly, some participants were concerned whether the system could handle the variability in the perspectives of a captured location in the Place channel. Participants liked that the system was able to recognize landmarks from a distance as that didn't require them "*to be up close and personal with [the] building*" (P8). However, they were frustrated when the system failed to recognize landmarks from a distance, which happened for a few participants due to variability of phone usage. Getting feedback at the right distance is important when the feedback is crucial to be received ahead of time (*e.g.*, detecting obstacles). Participants wanted to hear the feedback as "*further out it can tell*" —P10 or periodically when moving towards their target (*e.g.*, in decrements of "*50 feet, 25 feet, 10 feet*" — P10).

### Future Design Enhancements

Participants wanted access to more information with both more granularity and variety. For example, recognizing objects such as trash/recycling bins and drinking fountains, or landmarks such as pedestrian signals and restrooms. They wanted to identify objects that cannot be detected with their primary mobility aid such as railings when using a cane, empty tables when using a guide dog, and if there are people in the way when capturing a picture. In addition to the feedback about the environment, participants wanted precise directional guidance to reach their intended target as well as *in situ* guidance to use the system better. Precise directional guidance included providing information on the VI person's

spatial orientation, ego-centric directions, and distance from the object of interest. *In situ* guidance included: (i) how to hold the phone and manipulate the camera: "*I don't know if a sighted person would look down to find the table. So, does that mean I have to angle the phone down?*" (P5—congenitally blind participant); and (ii) identify and prompt the user when to use the system: "*I keep forgetting that sometimes there are signs that hang out perpendicular to the building; [...] signs are things that we avoid as blind people because they hurt.*" (P6). They also suggested using earcons to help them capture a picture better (*e.g.*, a beeping sound to guide the user in capturing the full sign). Additionally, participants mentioned varying directional instructions depending on an individual's residual vision, *e.g.*, using more visual instructions *vs.* more directional feedback. As P10 explains, "*I would give her [a friend with residual vision] more visual instructions because I know she can see that to a point.*" For a completely blind person, much more granular information is needed such as precise ego-centric directions to the object of interest (*e.g.*, "*men's bathroom 10 feet to your left*" or "*door 20 feet ahead*").

Finally, participants envisioned how such a vision-based system could be integrated or could work in tandem with other existing applications. Some examples included using it with native phone apps (*e.g.*, Photos), GPS-based apps such as Soundscape (*e.g.*, being able to set a beacon on the landmark of interest), using images from messaging applications or Google Street View as "places" to find, and mapping applications such as Google Maps: "*Collaboration is really an important thing when it comes to AI. If you could have the landmark feature integrated into [...] Google Maps for indoor navigation, that would be really nice in big hotels.*" (P11).

## DESIGN SPACE FOR LANDMARK-BASED SYSTEMS

As articulated by Williams *et al.* [58], individual differences play an important role in a VI person's use of navigation technology. Based on our studies' findings, literature on O&M training, and prior studies of VI peoples' navigation behaviors [2,8,46,58], we articulate a design space for creating adaptive systems using landmarks as the primary wayfinding strategy. Systems designed according to these principles would provide personalized information relevant to the user and the situational context. The need for tailored information based on the user's specific abilities is a key aspect in O&M training and the proposed principles strongly comply with the Ability-based Design paradigm [59].

As described earlier, landmarks have varied sensory properties such as having distinct colors, shapes, sizes, aromas, sounds, or textures. Landmark preferences depend on the landmark's availability, the mobility aid individuals use, residual vision, and the presence of other senses (*e.g.*, hearing). Based on these factors, the relevance of a particular landmark in a given situation may differ. We define a design space to capture this relationship by mapping a person's mobility need to the different affordances of a landmark and its environmental context. We break the design space into four components: (i) Visual Abilities, (ii) Mobility Aid, (iii) User Personality and Preferences, and (iv) Context.

### Visual Abilities

Adapting a system to VI user's visual abilities requires accommodating a person's use of their residual vision (if any) to navigate and how their navigation strategy impacts the choice of landmarks. During O&M training, an instructor assesses a user's residual vision to determine which landmarks would be usable. Relevant visual indicators include user's color perception, contrast sensitivity, visual acuity, and presence/absence of peripheral and central vision. As we saw from our study, landmarks are chosen based on their color, shape, size, and the location with respect to the user. For completely blind users, providing granular directional guidance is key. For people with residual vision, using visual instructions (*e.g.*, by describing visual features of the environment) is more appropriate. For example, for a person with color perception, an adaptive system should identify landmarks with distinct colors (*e.g.*, a bright red mailbox). Wayfinding apps could also be personalized in ways that best augments users' existing capabilities, *i.e.*, focusing only on calling out landmarks in the periphery if someone's central vision is intact. Alternatively, as suggested by one of our participants, a user may wish to specify in their profile that they would like an app to focus only on identifying landmarks in the region they can detect with their residual vision, so that they can then learn to attend to these landmarks in the future without the aid of the app.

### Mobility Aid

An adaptive system should consider the differences in the information received from a person's mobility aid. Mobility aids such as guide dogs and white canes have different affordances. For example, a guide dog is mainly used to avoid obstacles and is most suitable for search tasks, while a cane is effective for detecting obstacles and is most suitable for discovery tasks. These differences impact an individual's navigation strategy [58], as we saw VI individuals' ability to use our system differed depending on their primary mobility aid. For example, finding doorways is easier for guide dog users while it is a laborious and a manual process for cane users. On the other hand, guide dog users do not receive any tactile information of objects and surfaces around them. This suggests adaptive systems should make discovery of landmarks dynamic depending on a user's mobility aid [2,25]. For example, technology to assist in detecting tactile landmarks would be beneficial for guide dog users while systems that find structural landmarks such as doors and elevators would benefit cane users.

### User Personality and Preferences

An individual's confidence traveling independently is a major personality trait that influences how they wish to use guidance systems [2]. Confidence may depend on years of mobility training received and/or the number of years of sight loss. Such differences could inform the level of support and guidance needed from a wayfinding system. For example, a person with recent sight loss might need constant feedback

while a person who has never had vision may be more confident and may only need specific informational cues depending on what they want to achieve. In our study, we found that some participants were very task-oriented and only cared about the information relevant to the current context. In contrast, some participants wanted a full report on every landmark or object in the environment to get acclimatized and build a more complete mental model. Systems could support both pull and push interactions, allowing users to explicitly set their personal preferences.

## Context

Complementing prior work [1,2,34], the fourth aspect of adaptation relates to the contextual factors that determine landmark choices when on-the-go. Depending on a VI individual's situational context (*i.e.*, familiarity with the space, noise level, crowd density, weather, and time of day), the usefulness of a landmark will vary. For example, a user's familiarity with the location changes the navigation task from *discovery* (for unfamiliar spaces) to *search* (for familiar spaces). Certain landmark types may not be useful based on the environment (*e.g.*, sound landmarks when the environmental noise is high) or may not be available (*e.g.*, "ding" sounds if the elevator is out of service). When a location is crowded, navigation becomes slower and use of mobility aids becomes difficult (*e.g.*, guide dogs losing line of sight to targets such as doors); in such scenarios, detection of obstacles would be important to provide for a clear path to the user and their mobility aid. Finally, lighting conditions, depending on the time of day and weather, may affect computer vision technologies and users' residual vision.

## DISCUSSION

Using two exploratory studies, we investigated the characteristics of the last-few-meters wayfinding challenge and explored specific user needs in this space. From these studies, we articulated a design space for creating adaptive systems providing tailored feedback to VI pedestrians. This design space is not only limited to the last-few-meters problem but can also be applied to traditional point-to-point navigation applications where the primary means of navigation is walking.

In the last few meters, we found that the spatial relationship between the person and the surrounding landmarks and/or obstacles needs to be established (*e.g.*, elevator is 10 feet away from the user at their 2 o'clock). Amongst landmark categories, we found discovering structural landmarks was the most preferred across all participants. Usefulness of landmark categories depended on the user's situational context and personal preferences based on their vision level and mobility aid, and we captured this relationship in our proposed design space. Our findings demonstrate how Landmark AI can be useful in complementing a VI individual's mobility skills, *i.e.*, how the system would be used with their primary mobility aid and residual vision. We reflect on these findings and present implications for designing and developing camera-based AI tools for accessibility, and present limitations and future work.

## Implications for Camera Systems for VI Pedestrians

In this paper, we demonstrated the feasibility of using a smartphone camera-based system that provides near-real time information about the world within the accessibility context when the user is mobile. Within the three interaction paradigms we observed, *i.e.,* real-time scanning (Landmark channel), image capture (Signage channel), and hybrid—combining capturing images and real-time scanning (Place channel), participants preferred real-time scanning as it was fast, simple, and easy to use on-the-go. Capturing a good image was a frustrating experience [32,33]. Partial coverage or misfocus in capturing images of signs were common reasons for difficulty in using the channels. Applying blind photography principles [33] could help guide users to capture accurate pictures, though this remains a challenging area for further research. Additionally, participants preferred a simpler interaction than switching channels. Even though channels are an artifact of Seeing AI, this system design allowed us to analyze the implications and impact of these interaction paradigms: while channels simplify a system's technical implementation, they add overhead for the end user, and we recommend avoiding them.

Consistent with prior work [17,42], some participants had difficulty positioning the phone while walking. This caused misinterpretation of the app's feedback. Implementing camera guidance mechanisms [32,56] to handle hand-body coordination could resolve such difficulties. Alternatively, participants suggested using a wearable camera to allow hands free usage when they are mobile—critical for many participants due to either situational or motor impairments. Prior work [18] and industry solutions (*e.g.*, [84–86]) have looked into wearables for VI users. However, further work is required on wearable solutions to study scene capture accuracy and its impact on users' understanding and knowledge of the environment; Manduchi *et al.*'s investigation of blind users' ability to use smartphone object recognizers [42] is an important step in this direction.

## Implications for Vision Algorithms for Accessibility

*On-device Recognition*: In this paper, we looked at the application of deep neural networks (DNNs) for recognition tasks on mobile devices [28,31,63]. Use of fast and light-weight recognizers are crucial for providing real-time feedback when the user is mobile. We used a fast on-device recognizer based on SqueezeNet [31] to identify landmarks, making instantaneous response a possibility. However, a contributing factor to the signage channel being least preferred was the slow processing time due to dependence on cloud-based API calls. Current on-device recognizers lack the robustness in handling the variety of font styles encountered in the open world, particularly stylized logos common in commercial spaces. Future work from the vision community to develop on-device text recognition algorithms will be crucial in making signage recognition real-time. In addition to enabling real-time information, on-device recognition would also preserve privacy, especially for people captured in the scene.

***Need for Material Recognition***: Our design space highlights the importance of identifying a landmark's photometric and geometric properties to support varied vision levels in order to customize landmark detection. For this to happen, materials and texture recognition [5,12,29,50] would play a critical role, for example, detecting the material of the landmark and detecting changes in surface texture (for tactile landmarks). However, current computer vision algorithms [5,12,29] are not accurate enough, warranting an effort in improving their speed and accuracy. Combining material recognition with object recognition could also improve landmark recognition accuracy. In addition to materials, determining the color, shape, and size of landmarks is important when integrating them with object recognition.

***Implementing Place Recognition***: Landmark AI's place channel, which used a Wizard of Oz approach, was popular among study participants. Participants expressed interest in knowing whether the system would support recognizing the place if the original angle of capture differed from angle of the real-time feed. Prior work in robotics has looked into using deep learning approaches [11,41,53] and traditional computer vision techniques [14] for performing place recognition [24]. Future work in implementing a *real-time* place recognizer that is both *viewpoint invariant* and *time invariant* will be crucial in making this demonstrated experience a reality. Within the accessibility context, the place recognition problem can be constrained at two stages: (i) at the image capture stage, where unique landmarks are captured in the scene along with the location, and (ii) at the recognition stage, where performing a fuzzy match between the previously stored image and the current scene could be sufficient, thus circumventing the need for semantic scene understanding. This approach would be particularly useful for scenes for which specific classifiers have not been trained or that contain unusual uncommon objects.

***Achievable Accuracy****:* We found that participants preferred certain types of landmarks such as doors over others. This suggests that we may not need general-purpose recognizers that classify a wide range of possible objects, a daunting task for current computer vision algorithms. Instead, collecting a large and realistic dataset of common landmarks and objects (*e.g.*, doors of different types), combined with counterexamples of objects that are similar and confusable with the object of interest (*e.g.*, full-pane windows) would be a priority. Building a robust recognition model for a smaller (but important) set of objects could have a significant impact on VI users' daily navigation abilities. Our design decision of using simpler vision models with preset landmarks was guided by this fact to maintain a reasonable level of accuracy.

In our system, we cared more about precision (low false positives) than recall (low false negatives). Ideally, there should be a balance between the two. However, realistically there are high chances of the results being skewed. In those cases, low precision causes more harm than low recall. In our study, we found participants getting frustrated with false positives, making it hard to rely on the system. Participants did understand that a system cannot be perfect, and they valued access to rich contextual information. However, the system cannot "*provide the user too much of wrong information, because that will directly confuse the user more than really help them out.*" (P9). DNNs have been found to get easily "fooled" even with a high confidence threshold [44]. For a deployable level of accuracy, using computer vision techniques alone may be insufficient. Potential solutions relying on humans to use their own judgment to reason about the inference (*e.g.*, using explainable AI techniques [26]) or using heuristics and sensor fusion techniques to supplement the vision results could help establish more confidence in AI-based navigation aids.

## Limitations and Future Work
Two main limitations may impact our findings. First, due to Landmark AI's high rate of false positives, participants were often frustrated and confused. While we believe that accuracy should have been better, this allowed us to understand the implications of poor accuracy, a likely scenario in the open world in the near-term with state-of-the-art AI. Studying how people learn to adapt to system inaccuracies will be valuable for understanding usage of fault-prone AI systems [1]. Second, Landmark AI did not provide navigational guidance to reach the landmark target once it was identified, an important characteristic for a wayfinding system [42,60]. However, this gave us an opportunity to investigate the guidance needs in the last few meters. Indeed, we observed that the system does not have to be hyper-accurate with such guidance, as one's existing mobility skills (through O&M training or otherwise) plays an important role of being independent. As participant P10 summarizes, "*At some point, you got to leave something out to the user to use their brain. Some people want to be spoon-fed every single little bit of information, but how do you learn if you don't find the stuff out for yourself?*".

## CONCLUSION
In this paper, we investigated the last-few-meters wayfinding problem. Our formative study identified common challenges faced in the last few meters, how VI users currently overcome them, and where current technology falls short. Based on these findings, we designed *Landmark AI* and elicited feedback on the usefulness of the system design via a design probe study. Using qualitative data analysis, we found that an individual's choice of mobility aid (*e.g.*, guide dogs or white canes) and their visual ability impacted the manner in which they used the system and the provided feedback. We captured this rich relationship between the information types and an individual's mobility needs in a design space for creating adaptive systems and presented a set of design implications for future camera-based AI systems for people with VI.

**REFERENCES**

[1] Ali Abdolrahmani, William Easley, Michele Williams, Stacy Branham, and Amy Hurst. 2017. Embracing Errors: Examining How Context of Use Impacts Blind Individuals' Acceptance of Navigation Aid Errors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4158–4169. https://doi.org/10.1145/3025453.3025528

[2] Dragan Ahmetovic, João Guerreiro, Eshed Ohn-Bar, Kris M. Kitani, and Chieko Asakawa. 2019. Impact of Expertise on Interaction Preferences for Navigation Assistance of Visually Impaired Individuals. In *Web for All Conference (W4A)*.

[3] Mauro Avila and Limin Zeng. 2017. A Survey of Outdoor Travel for Visually Impaired People Who Live in Latin-American Region. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '17*, 9–12. https://doi.org/10.1145/3056540.3064953

[4] Yicheng Bai, Wenyan Jia, Hong Zhang, Zhi-Hong Mao, and Mingui Sun. 2014. Landmark-Based Indoor Positioning for Visually Impaired Individuals. *International conference on signal processing proceedings. International Conference on Signal Processing* 2014: 678–681. https://doi.org/10.1109/ICOSP.2014.7015087

[5] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. 2015. Material Recognition in the Wild with the Materials in Context Database. *Computer Vision and Pattern Recognition (CVPR)*.

[6] Jeffrey P. Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010. VizWiz::LocateIt - Enabling blind people to locate objects in their environment. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, 65–72. https://doi.org/10.1109/CVPRW.2010.5543821

[7] Jeffrey P. Bigham, Samual White, Tom Yeh, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, and Brandyn White. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*, 333. https://doi.org/10.1145/1866029.1866080

[8] Bruce B Blasch, William R Wiener, and Richard L Welsh. 1997. Foundations of orientation and mobility.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2: 77–101.

[10] Long Chen, Bao-long Guo, and Wei Sun. 2010. Obstacle detection system for visually impaired people based on stereo vision. In *2010 Fourth International Conference on Genetic and Evolutionary Computing*, 723–726.

[11] Zetao Chen, Adam Jacobson, Niko Sunderhauf, Ben Upcroft, Lingqiao Liu, Chunhua Shen, Ian Reid, and Michael Milford. 2017. Deep Learning Features at Scale for Visual Place Recognition.

[12] Mircea Cimpoi, Subhransu Maji, and Andrea Vedaldi. 2015. Deep filter banks for texture recognition and segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3828–3836.

[13] James Coughlan, R Manduchi, and Huiying Shen. 2006. Cell Phone-based Wayfinding for the Visually Impaired. In *First International Workshop on Mobile Vision*.

[14] Mark Cummins and Paul Newman. 2008. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research* 27, 6: 647–665. https://doi.org/10.1177/0278364908090961

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[16] Mary Beatrice Dias, Ermine A Teves, George J Zimmerman, Hend K Gedawy, Sarah M Belousov, and M Bernardine Dias. 2015. Indoor Navigation Challenges for Visually Impaired People. In *Indoor Wayfinding and Navigation*, Hassan A Karimi (ed.). 141–164.

[17] Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. 2012. The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 425. https://doi.org/10.1145/2207676.2207735

[18] Alexander Fiannaca, Ilias Apostolopoulous, and Eelke Folmer. 2014. Headlock: A Wearable Navigation Aid That Helps Blind Cane Users Traverse Large Open Spaces. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '14), 19–26. https://doi.org/10.1145/2661334.2661453

[19] John C. Flanagan. 1954. The critical incident technique. *Psychological Bulletin* 51, 4: 327–358. https://doi.org/10.1037/h0061470

[20] Aura Ganz, Siddhesh Rajan Gandhi, James Schafer, Tushar Singh, Elaine Puleo, Gary Mullett, and Carole Wilson. 2011. PERCEPT: Indoor navigation for the blind and visually impaired. In *2011 Annual*

*International Conference of the IEEE Engineering in Medicine and Biology Society*, 856–859. https://doi.org/10.1109/IEMBS.2011.6090223

[21] Aura Ganz, James M. Schafer, Yang Tao, Carole Wilson, and Meg Robertson. 2014. PERCEPT-II: Smartphone based indoor navigation system for the blind. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3662–3665. https://doi.org/10.1109/EMBC.2014.6944417

[22] Cole Gleason, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2018. FootNotes: Geo-referenced Audio Annotations for Nonvisual Exploration. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3: 1–24. https://doi.org/10.1145/3264919

[23] Vladimir Golkov, Alexey Dosovitskiy, Jonathan I. Sperl, Marion I. Menzel, Michael Czisch, Philipp Sämann, Thomas Brox, and Daniel Cremers. 2016. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. *IEEE Transactions on Medical Imaging* 35, 5: 1344–1351. https://doi.org/10.1109/TMI.2016.2551324

[24] Reginald G Golledge. 2004. Place Recognition and Wayfinding : Making Sense of Space.

[25] João Guerreiro, Eshed Ohn-Bar, Dragan Ahmetovic, Kris Kitani, and Chieko Asakawa. 2018. How Context and User Behavior Affect Indoor Navigation Assistance for Blind People. In *Proceedings of the Internet of Accessible Things on - W4A '18*, 1–4. https://doi.org/10.1145/3192714.3192829

[26] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), and Web.*

[27] Nghia Ho and Ray Jarvis. 2008. Towards a platform independent real-time panoramic vision based localisation system. In *Australasian Conference on Robotics and Automation.*

[28] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR* abs/1704.0.

[29] Diane Hu, Liefeng Bo, and Xiaofeng Ren. 2011. Toward Robust Material Recognition for Everyday Objects. In *BMVC*, 6.

[30] Feng Hu, Zhigang Zhu, and Jianting Zhang. 2014. Mobile Panoramic Vision for Assisting the Blind via Indexing and Localization. In *European Conference on Computer Vision*, 600–614.

[31] Forrest N Iandola, Matthew W Moskewicz, Khalid Ashraf, Song Han, William J Dally, and Kurt Keutzer. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *CoRR* abs/1602.0.

[32] Chandrika Jayant and Chandrika. 2010. MobileAccessibility: camera focalization for blind and low-vision users on the go. *ACM SIGACCESS Accessibility and Computing*, 96: 37–40. https://doi.org/10.1145/1731849.1731856

[33] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11*, 203. https://doi.org/10.1145/2049536.2049573

[34] Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Valeria Alampi, Hironobu Takagi, and Chieko Asakawa. 2018. Insights on Assistive Orientation and Mobility of People with Visual Impairment Based on Large-Scale Longitudinal Data. *ACM Transactions on Accessible Computing* 11, 1: 1–28. https://doi.org/10.1145/3178853

[35] Shaun K. Kane, Brian Frey, and Jacob O. Wobbrock. 2013. Access lens: a gesture-based screen reader for real-world documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 347. https://doi.org/10.1145/2470654.2470704

[36] J. F. Kelley and J. F. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems* 2, 1: 26–41. https://doi.org/10.1145/357417.357420

[37] Eunjeong Ko, Eun Kim, Eunjeong Ko, and Eun Yi Kim. 2017. A Vision-Based Wayfinding System for Visually Impaired People Using Situation Awareness and Activity-Based Instructions. *Sensors* 17, 8: 1882. https://doi.org/10.3390/s17081882

[38] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction.*

[39] Jacobus Cornelius Lock, Grzegorz Cielniak, and Nicola Bellotto. 2017. A Portable Navigation System with an Adaptive Multimodal Interface for the Blind. In *2017 AAAI Spring Symposium Series.*

[40] Jack M Loomis, Roberta L Klatzky, Reginald G Golledge, and others. 2001. Navigating without vision: basic and applied research. *Optometry and vision science* 78, 5: 282–289.

[41] Stephanie Lowry, Niko Sunderhauf, Paul Newman, John J. Leonard, David Cox, Peter Corke, and Michael J. Milford. 2016. Visual Place Recognition: A Survey. *IEEE Transactions on Robotics* 32, 1: 1–19. https://doi.org/10.1109/TRO.2015.2496823

[42] Roberto Manduchi and James M. Coughlan. 2014. The last meter: blind visual guidance to a target. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 3113–3122. https://doi.org/10.1145/2556288.2557328

[43] Pratap Misra and Per Enge. 2006. Global Positioning System: signals, measurements and performance second edition. *Massachusetts: Ganga-Jamuna Press*.

[44] Anh Nguyen, Jason Yosinski, and Jeff Clune. 2015. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In *Computer Vision and Pattern Recognition (CVPR '15)*.

[45] J. Eduardo Pérez, Myriam Arrue, Masatomo Kobayashi, Hironobu Takagi, and Chieko Asakawa. 2017. Assessment of Semantic Taxonomies for Blind Indoor Navigation Based on a Shopping Center Use Case. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work - W4A '17*, 1–4. https://doi.org/10.1145/3058555.3058575

[46] Paul E Ponchillia and Susan Kay Vlahas Ponchillia. 1996. *Foundations of rehabilitation teaching with persons who are blind or visually impaired*. American Foundation for the Blind.

[47] Jose Rivera-Rubio, Saad Idrees, Ioannis Alexiou, Lucas Hadjilucas, and Anil A Bharath. 2013. Mobile Visual Assistive Apps: Benchmarks of Vision Algorithm Performance. In *New Trends in Image Analysis and Processing -- ICIAP 2013*, 30–40.

[48] Mary Beth Rosson and John M Carroll. 2009. Scenario based design. *Human-computer interaction. boca raton, FL*: 145–162.

[49] Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. NavCog3: An Evaluation of a Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17*, 270–279. https://doi.org/10.1145/3132525.3132535

[50] Gabriel Schwartz and Ko Nishino. 2018. Recognizing Material Properties from Images.

[51] M Serrão, João M F Rodrigues, José I Rodrigues, and J M Hans du Buf. 2012. Indoor localization and navigation for blind persons using visual landmarks and a GIS. *Procedia Computer Science* 14: 65–73.

[52] Huiying Shen and James M Coughlan. 2012. Towards a real-time system for finding and reading signs for visually impaired users. In *International Conference on Computers for Handicapped Persons*, 41–47.

[53] Niko Sünderhauf, Sareh Shirazi, Adam Jacobson, Feras Dayoub, Edward Pepperell, Ben Upcroft, and Michael Milford. 2015. Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free. *ARC Centre of Excellence for Robotic Vision; Science & Engineering Faculty*.

[54] Yingli Tian, Xiaodong Yang, and Aries Arditi. 2010. Computer vision-based door detection for accessibility of unfamiliar environments to blind persons. In *International Conference on Computers for Handicapped Persons*, 263–270.

[55] Sylvie Treuillet, Eric Royer, Thierry Chateau, Michel Dhome, Jean-Marc Lavest, and others. 2007. Body Mounted Vision System for Visually Impaired Outdoor and Indoor Wayfinding Assistance. In *CVHI*.

[56] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*, 95. https://doi.org/10.1145/2384916.2384934

[57] Anthony J Viera, Joanne M Garrett, and others. 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med* 37, 5: 360–363.

[58] Michele A. Williams, Amy Hurst, and Shaun K. Kane. 2013. "Pray before you step out": describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '13*, 1–8. https://doi.org/10.1145/2513383.2513449

[59] Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-Based Design: Concept, Principles and Examples. *ACM Transactions on Accessible Computing* 3, 3: 1–27. https://doi.org/10.1145/1952383.1952384

[60] Zhuorui Yang and Aura Ganz. 2017. Egocentric Landmark-Based Indoor Guidance System for the Visually Impaired. *International Journal of E-Health and Medical Communications* 8, 3: 55–69. https://doi.org/10.4018/IJEHMC.2017070104

[61] Hanlu Ye, Meethu Malu, Uran Oh, Leah Findlater, Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. 2014. Current and future mobile and wearable device use by people with visual impairments. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 3123–3132. https://doi.org/10.1145/2556288.2557085

[62] Limin Zeng. 2015. A survey: outdoor mobility experiences by the visually impaired. *Mensch und Computer 2015--Workshopband*.

[63] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2017. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *CoRR* abs/1707.0.

[64] Yu Zhong, Pierre J. Garrigues, and Jeffrey P. Bigham. 2013. Real time object scanning using a mobile phone and cloud-based visual search engine. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '13*, 1–8. https://doi.org/10.1145/2513383.2513443

[65] BlindSquare. Retrieved September 18, 2018 from http://www.blindsquare.com/

[66] Microsoft Soundscape | A map delivered in 3D sound. Retrieved September 15, 2018 from https://www.microsoft.com/en-us/research/product/soundscape/

[67] The Seeing Eye GPS^TM App for cell-enabled iOS devices. Retrieved September 18, 2018 from http://www.senderogroup.com/products/shopseeingeye gps.html

[68] Autour. Retrieved September 18, 2018 from http://autour.mcgill.ca/en/

[69] GPS.gov: GPS Accuracy. Retrieved September 17, 2018 from https://www.gps.gov/systems/gps/performance/accuracy/

[70] Indoor Movement and Orientation: Use Your Senses | VisionAware. Retrieved September 17, 2018 from http://www.visionaware.org/info/essential-skills-2/an-introduction-to-orientation-and-mobility-skills/indoor-movement-and-orientation-with-vision-impairment/235

[71] PERCEPT: Indoor Wayfinding for The Blind and Visually Impaired. Retrieved September 16, 2018 from http://www.perceptwayfinding.com/

[72] NAVATAR- A low cost indoor navigation system for blind students. Retrieved September 14, 2018 from https://humanpluslab.github.io/Navatar/

[73] BlindWays - free app makes using public transit easier for riders with visual impairments | Perkins School for the Blind. Retrieved November 15, 2018 from https://www.perkins.org/access/inclusive-design/blindways

[74] Seeing AI | Talking camera app for those with a visual impairment. Retrieved September 15, 2018 from https://www.microsoft.com/en-us/seeing-ai

[75] Aipoly Vision: Sight for Blind & Visually Impaired. Retrieved from https://itunes.apple.com/us/app/aipoly-vision-sight-for-blind-visually-impaired/id1069166437?mt=8

[76] KNFB Reader. Retrieved October 9, 2018 from https://knfbreader.com/

[77] Be My Eyes. Retrieved October 9, 2018 from https://www.bemyeyes.com/

[78] BeSpecular. Retrieved October 9, 2018 from https://www.bespecular.com/

[79] Aira. Retrieved June 26, 2018 from https://aira.io/

[80] TapTapSee - Blind and Visually Impaired Assistive Technology. Retrieved October 8, 2018 from https://taptapseeapp.com/

[81] CloudSight AI. Retrieved October 8, 2018 from https://cloudsight.ai/

[82] LookTel Recognizer for iPhone. Retrieved July 18, 2019 from http://www.looktel.com/recognizer

[83] Cognitive Services | Microsoft Azure. Retrieved September 19, 2018 from https://azure.microsoft.com/en-us/services/cognitive-services/

[84] Introducing Lookout: helping users who are blind learn about the world around them - Google Accessibility Blog – Google Accessibility. Retrieved November 9, 2018 from https://www.google.com/accessibility/blog/post/announce-lookout.html

[85] With Lookout, discover your surroundings with the help of AI. Retrieved October 30, 2018 from https://www.blog.google/outreach-initiatives/accessibility/lookout-discover-your-surroundings-help-ai/

[86] IRSAW: Intel RealSense Spatial Awareness Wearable. Retrieved April 28, 2019 from https://github.com/IRSAW/IRSAW

[87] 2017. WHO | Global data on visual impairment. *WHO*.