Towards Street Camera-based Outdoor Navigation for Blind Pedestrians

Gaurav Jain
Columbia University
New York, NY, USA

Basel Hindi
Columbia University
New York, NY, USA

Mingyu Xie
Columbia University
New York, NY, USA

Zihao Zhang
Columbia University
New York, NY, USA

Koushik Srinivasula
Columbia University
New York, NY, USA

Mahshid Ghasemi
Columbia University
New York, NY, USA

Daniel Weiner*
Lehman College
New York, NY, USA

Xin Yi Therese Xu*
Pomona College
Claremont, CA, USA

Sophie Ana Paris*
New York University
New York, NY, USA

Chloe Tedjo*
Texas A&M University
College Station, TX, USA

Josh Bassin*
The Penn State University
State College, PA, USA

Michael Malcolm*
SUNY at Albany
Albany, NY, USA

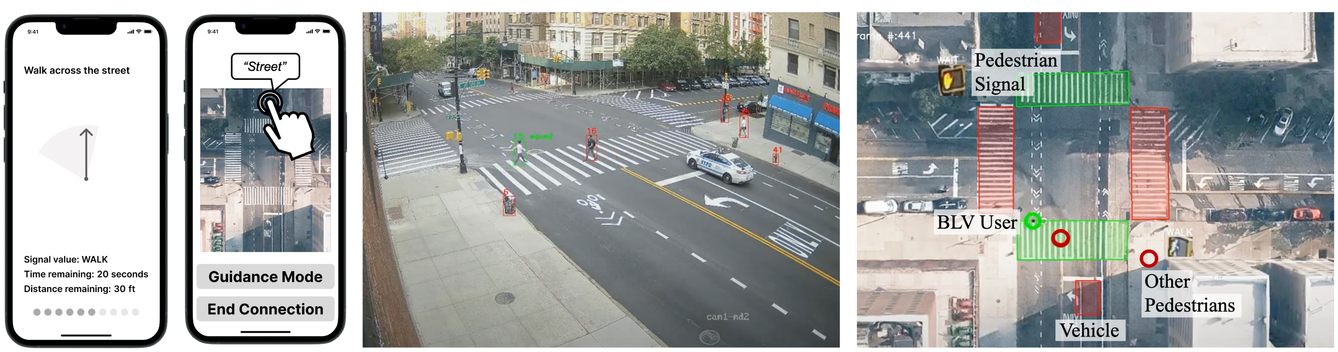
Mehmet Turkcan
Columbia University
New York, NY, USA

Javad Ghaderi
Columbia University
New York, NY, USA

Zoran Kostic
Columbia University
New York, NY, USA

Gil Zussman
Columbia University
New York, NY, USA

Brian A. Smith
Columbia University
New York, NY, USA



(a) Guidance (b) Exploration (c) Street Camera Video Feed (d) Bird's Eye View Map

Figure 1: Overview of the street camera-based navigation system. The system provides real-time auditory feedback to help BLV users avoid obstacles, know exactly when to cross the street, and understand the overall layout of the environment. Blind and low-vision (BLV) pedestrians interact with the system via a companion smartphone app that offers two navigation modes: (a) guidance mode and (b) exploration mode. The system leverages street cameras to enable these two navigation modes, by first processing the (c) COSMOS testbed's street camera video feed using computer vision to identify user's position, nearby obstacles, and pedestrian signals; and then transforming it onto a (d) bird's eye view map for effectively conveying instructions.

ABSTRACT

Blind and low-vision (BLV) people use GPS-based systems for outdoor navigation assistance, which provide instructions to get from one place to another. However, such systems do not provide users with real-time, precise information about their location and surroundings which is crucial for safe navigation. In this work, we investigate whether street cameras can be used to address aspects of navigation that BLV people still find challenging with existing GPS-based assistive technologies. We conducted formative interviews with six BLV participants to identify specific challenges they face in outdoor navigation. We discovered three main challenges:

*This work was done while Daniel Weiner, Xin Yi Therese Xu, Sophie Ana Paris, Chloe Tedjo, Josh Bassin, and Michael Malcolm were interns at Columbia University.

ASSETS '23, October 22–25, 2023, New York, NY, USA

ACM ISBN 979-8-4007-0220-4/23/10.
https://doi.org/10.1145/3597638.3614498

anticipating environment layouts, avoiding obstacles while following directions, and crossing noisy street intersections. To address these challenges, we are currently developing a street camera-based navigation system that provides real-time auditory feedback to help BLV users avoid obstacles, know exactly when to cross the street, and understand the overall layout of the environment. We close by discussing our evaluation plan.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility systems and tools**.

KEYWORDS

Visual impairments, outdoor navigation, street camera, computer vision, testbed evaluation

ACM Reference Format:

Gaurav Jain, Basel Hindi, Mingyu Xie, Zihao Zhang, Koushik Srinivasula, Mahshid Ghasemi, Daniel Weiner, Xin Yi Therese Xu, Sophie Ana Paris, Chloe Tedjo, Josh Bassin, Michael Malcolm, Mehmet Turkcan, Javad Ghaderi, Zoran Kostic, Gil Zussman, and Brian A. Smith. 2023. Towards Street Camera-based Outdoor Navigation for Blind Pedestrians. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23), October 22–25, 2023, New York, NY, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3597638.3614498

1 INTRODUCTION

Outdoor navigation in unfamiliar environments is challenging for blind and low-vision (BLV) people. GPS-based assistive technologies, such as BlindSquare [25] and Microsoft Soundscape [15], are commonly used by BLV people to receive guidance instructions to a chosen point of interest (POI). While GPS-based systems provide information about route to the destination, they fail to assist with other aspects of outdoor navigation that require real-time and precise knowledge of the user's location and surroundings. For instance, BLV people face difficulties in avoiding obstacles (e.g., other pedestrians, vehicles) [29, 31], maintaining a straight path [28], and crossing street intersections [2, 13, 24]. Thus, there is a need to investigate alternate technologies that can support the precise and real-time aspects of BLV pedestrians' outdoor navigation.

One particularly promising alternative is to leverage already-instrumented street cameras in outdoor environments, that are increasingly being installed in cities for public safety, surveillance, and traffic management-related applications [3, 5, 10, 19, 23]. While accessibility is not street cameras' primary purpose, they have the potential to be repurposed as navigation assistance systems.

In this work, we investigate street cameras' potential for supporting aspects of outdoor navigation that require precise and real-time knowledge of BLV pedestrians' location and surroundings. To this end, we take preliminary steps to answer the following research questions:

RQ1. What aspects of outdoor navigation do BLV people find challenging when using GPS-based assistive technology?

RQ2. How should street camera-based systems be designed to address these challenging aspects of outdoor navigation?

RQ3. To what extent do street camera-based navigation systems address these outdoor navigation challenges?

To answer RQ1, we performed formative interviews with six BLV participants and found that anticipating environment layouts, avoiding obstacles while following directions, and crossing street intersections in noisy environments are challenging aspects of outdoor navigation that GPS-based systems fail to address.

To answer RQ2, we are currently prototyping a street camera-based navigation system that addresses the challenges revealed in RQ1. To interact with the street camera system, BLV pedestrians use a companion smartphone application and Bluetooth earpiece. When navigating outdoors, BLV users simply wave one hand over their head and the street camera system embedded within the environment recognizes their precise location on the street using computer vision. Once localized, pedestrians can choose to receive turn-by-turn instructions to a nearby POI or explore the layout of the environment. As users navigate through the environment, they receive real-time auditory feedback via the Bluetooth earpiece that helps prevent veering off the path, avoid obstacles, and know exactly when to cross the street; as shown in Figure 1. We deployed our system in the NSF PAWR COSMOS wireless edge-cloud testbed in New York City (NYC) [34, 42]. In particular, we used one of COSMOS' street-level cameras deployed on the second floor of Columbia's Mudd building and one of COSMOS' edge servers as the computational server.

We close by discussing our plan for evaluating the street camera-based navigation system to answer RQ3.

2 RELATED WORK

Existing approaches for outdoor navigation primarily rely on GPS-based navigation systems for providing turn-by-turn instructions and information about nearby POIs [15, 17, 25]. The GPS signal, however, offers poor precision with localization errors as big as tens of meters [1, 26, 43]. The accuracy is lower in densely populated cities [40], which is even more concerning given that a disproportionately high percentage of BLV people live in cities [14]. Despite GPS-based systems' undeniable impact on helping BLV people in outdoor navigation, their low precision and inability to provide real-time support for avoiding obstacles and veering off the path limits their usability as a standalone navigation solution. Our work attempts to investigate street cameras' potential as an alternative solution for providing precise and real-time navigation assistance along with turn-by-turn guidance. Prior work has explored the use of surveillance cameras in indoor environments for robot navigation [4, 27, 30, 36]. Our work focuses on leveraging street cameras in outdoor environments as an assistive navigation tool for people with visual impairments.

Another common approach for outdoor navigation is to develop personalized, purpose-built, assistive devices that support with either crossing streets [13, 21, 37], recording routes [43], or avoiding obstacles [7, 8, 18, 22, 33, 41]. While these solutions address some of the precise and real-time aspects of BLV people's outdoor navigation, support for turn-by-turn navigation is missing. Furthermore, these systems place the burden of purchasing costly devices onto the BLV users. Our work, by contrast, explores the possibility of using street cameras that already exist in an environment to provide comprehensive outdoor navigation assistance. We investigate repurposing existing hardware in outdoor environments to support

Table 1: Self-reported demographics of our participants. Gender information was collected as a free response; our participants identified themselves as female (F) or male (M). Participants rated their assistive technology (AT) familiarity on a scale of 1–5.

PID	Gender	Age	Race	Occupation	Vision ability	Onset	Mobility aid	AT familiarity (1–5)
P1	F	29	White	Claims expert	Totally blind	Birth	White cane	3: Moderately familiar
P2	F	61	White	Retired	Light perception only	Age 6	Guide dog	1: Not at all familiar
P3	F	66	White	Retired	Totally blind	Age 58	Guide dog	2: Slightly familiar
P4	M	48	Black	Unemployed	Light perception only	Age 32	White cane	3: Moderately familiar
P5	M	27	White/Asian	Unemployed	Totally blind	Birth	White cane	3: Moderately familiar
P6	M	38	White	AT instructor	Totally blind	Birth	White cane	5: Extremely familiar

accessibility applications, thus imbuing accessibility within the city infrastructure directly at no additional cost to the BLV user.

3 FORMATIVE INTERVIEWS

We conducted semi-structured interviews with six BLV participants to answer RQ1: *What aspects of outdoor navigation do BLV people find challenging when using GPS-based assistive technology?*

3.1 Methods

We recruited six BLV participants (three males and three females; aged 29–66) by posting on social media platforms and snowball sampling [11]. Table 1 summarises the participants' information. All interviews were conducted over Zoom and lasted about 90 minutes. Participants were compensated $25 for this IRB approved study.

To understand the specific aspects of outdoor navigation that BLV people find challenging, we used a recent critical incident technique (CIT) [9], in which we asked participants to recall and describe a recent time when they navigated outdoor environments using GPS-based assistive technology (AT). For example, we first asked participants to name the AT they commonly use and then asked them to elaborate on their recent experience of using it: *"So, you mentioned using BlindSquare a lot. When was the last time you used it?"* Then, we initiated a discussion by establishing the scenario for them: *"Now, let's walk through your visit from the office to this restaurant. Suppose, I spotted you at your office. What would I observe? Let's start with you getting out of your office building."* We asked follow-up questions to gain insights into what made the aspects of outdoor navigation challenging and what additional information could help address them.

3.2 Findings: Challenging Aspects of Outdoor Navigation

We found three major themes for aspects of outdoor navigation that participants found challenging when using GPS-based AT.

3.2.1 Anticipating environment layout. GPS-based systems, like BlindSquare [25], offer navigation instructions that follow a direct path to the destination, often referred to as "as the crow flies," rather than providing detailed instructions through a poly-line path that guide BLV people as per the environment layout. Since *"not everything is organized in the ideal grid-like way"* (P1), participants reported facing difficulties in following these instructions, failing to confidently act upon the instructions without any knowledge of the shape and layout of the environment. P3 recalls: *"I didn't know if crosswalks were straight or curved or if they were angled. [It was hard] to figure out which way you needed to be to be in the crosswalk."* Many participants cited problems such as making the wrong turns into unexpected "alleyways" (P1, P2, P4) that landed them in dangerous situations with "cars coming through" (P2). Unfamiliar layouts also caused participants to veer off the sidewalks and end up in streets.

3.2.2 Avoiding obstacles while following instructions. Participants reported using their existing mobility aids along with GPS-based systems for getting directions. While doing so, participants found it challenging to keep their concentration on identifying obstacles and often bumped into things that they would have otherwise identified via their white cane. P2 shared an instance where *"there were traffic cones [and] I tripped over those"* while following directions. Both dynamic obstacles (e.g., other pedestrians, cars) and temporarily placed stationary obstacles (e.g., triangle sandwich board sign –P3) were hard to navigate around. P4 echoed this sentiment: *"You know how many times I've walked into the sides of cars even though I have the right of way. Drivers have gotten angry, accusing me of scratching their vehicles. It can spoil your day [and make] you feel insecure."*

3.2.3 Crossing street intersections safely. In line with prior research [13], our participants expressed that crossing streets was still a major challenge for them. Most participants mentioned relying on audio cues to identify the flow of traffic, but found it to be often insufficient in practice: *"yeah, it can be tricky, because [there may be] really loud construction nearby that can definitely throw me off because I'm trying to listen to the traffic"* (P1). Furthermore, not knowing the duration of the signals and the length of the crosswalk affected their confidence as they feared getting in trouble: *"I don't want to be caught in the middle [of the street]"* (P4).

4 STREET CAMERA-BASED NAVIGATION SYSTEM

In this section, we introduce a navigation system that we are currently developing to answer RQ2: *How should street camera-based systems be designed to address the challenging aspects of outdoor navigation?* The system consists of three components: *(i)* street camera, *(ii)* computational server, and *(iii)* smartphone app. These components interact with each other to facilitate two navigation modes that together address BLV people's challenges to outdoor navigation, which we discovered in our formative interviews.

4.1 System Components

The system consists of three main components: (i) street camera, (ii) computational server, and (iii) smartphone application.

Street camera. The system uses a street-level view camera which is part of the NSF PAWR COSMOS wireless edge-cloud testbed in NYC [34, 42]. The camera is mounted at the second floor of the Columbia Mudd building, viewing a four-way street intersection at 120th St. and Amsterdam Ave., NYC, as shown in Figure 1c. Anonymized video samples (without special gestures) from this camera can be found online [6]. The video feed from the camera is directly streamed onto one of COSMOS' computational servers for processing.

Computational server. The computational server processes the video feed using GStreamer-based pipelines [38] equipped with the YOLOv8 model [39] as an object detector and the Nvidia DCF-based tracker to track pedestrians and vehicles and identify pedestrian signals. Using the camera view and a corresponding image from Apple Maps' street view of the same intersection, the system finds visual correspondences to generate a bird's-eye view representation of the environment (Figure 1d). Additionally, it stores the map information that includes labeled regions (e.g., streets, crosswalks, sidewalks, pedestrian lights) and the location of relevant POIs (e.g., pharmacy, café) within the bird's-eye view representation. Similar to prior work in indoor navigation [1, 12, 35], the map information is prepared manually by an administrator and loaded onto the server beforehand.

Smartphone application. Figure 1a–b shows the iOS app that acts as an interface between the user and the computational server, enabling them to access the map information and to receive real-time audio feedback via a Bluetooth earpiece. To alleviate concerns around revealing private identifiable information from the video feeds (e.g., pedestrian's faces and vehicle's license plate), the server only sends processed information such as navigation instructions, positions and generic labels of obstacles (e.g., "vehicle" at 2 o'clock) to the smartphone app instead of the video itself.

4.2 User Interaction

BLV pedestrians use the smartphone app to establish a connection with the server via a localization mechanism. Once localized, users are offered two navigation modes: (i) guidance mode, and (ii) exploration mode.

Pedestrian Localization. To determine the user's position on the bird's-eye view map, the system must differentiate them from other pedestrians on the street. We achieved this by introducing an action recognition module that can identify users from the second floor camera feed. The smartphone app asks the user to initialize the system with their current position by simply waving one hand above their head for a few seconds, which is detected by the action recognition module. We chose this action based on discussions with several BLV individuals and most agreed that this single-handed action was both convenient and socially acceptable to them. Internally, the action recognition module is implemented as a CLIP model [32] that computes visual similarity of each detected pedestrian from the second floor camera with the following language prompts: "*person walking*" and "*person waving hand*." We experimentally fine-tuned the confidence thresholds.

Navigation modes. To address the challenging aspects of outdoor navigation that we identified from our formative study, we designed the street camera-based navigation system to support the following two modes of navigation:

(i) Guidance Mode. Figure 1a shows this mode, where BLV users can choose a destination from the list of nearby POIs and receive real-time audio feedback in the form of turn-by-turn instructions. Similar to prior work in indoor navigation [1], we represent the birds-eye view map as a graph representation consisting of POIs and street corners as nodes that act as way-points. The knowledge of the user's precise position enables the system to provide audio cues that help prevent veering off the path between way-points (Section 3.2.1).

To address BLV users' challenges to avoid obstacles while following instructions (Section 3.2.2), the street camera-based system notifies users of obstacles —both moving and fixed— by specifying their relative spatial location and category (e.g., by announcing "*pedestrian at 2'o clock, 5 feet away*"). Our current implementation offers support for dynamic obstacles such as pedestrians and vehicles, along with fixed ones such as poles, trashcans, and parked vehicles. Internally, we implement this by tracking all these elements within the space and predicting positional overlaps in bird's-eye view. For dynamic obstacles, specifically vehicles, we plan on adapting our prediction module to also account for their speed.

To address BLV people's challenges in crossing street intersections safely (Section 3.2.3), the system dynamically updates the internal graph representation to temporarily remove crosswalks that have pedestrian signals reading "wait" and reinstates it when they read: "walk.'' Once the system reinstates the crosswalk, it provides users precise information about the time remaining to cross and the distance to the other end of the crosswalk. The system gathers this information by first automatically detecting the signal state (i.e., walk vs. wait) and then computing the time it takes to change over a complete cycle.

(ii) Exploration Mode. Figure 1b shows this mode, where BLV users can choose to navigate the environment without any specific destination in mind. Similar to guidance mode, this mode also provides users real-time feedback to prevent veering (Section 3.2.1), avoid obstacles (Section 3.2.2), and cross street intersections safely (Section 3.2.3). Additionally, this mode is designed to address BLV users' challenge to anticipate environment layouts (Section 3.2.1). The user can scrub their finger on the smartphone to learn (via haptic feedback) the bird's-eye view map's shape and layout, which has been found to provide BLV people spatial understanding of the environment [16]. Prior work on image accessibility also shows that direct manipulation via touchscreen-based interfaces helps BLV users effectively explore images [20]. Our current implementation allows users to move their finger across the map on the smartphone app, reading out the corresponding region labels (e.g., street, crosswalk, sidewalk). We plan on extending this touchscreen-based exploration tool to also convey users' current position and POIs.

5 EVALUATION PLAN

We will conduct user studies with BLV pedestrians to evaluate the street camera-based navigation system (i.e., to answer RQ3). In this study, we will compare participants' experience of navigating street intersections using the proposed system and a commonly used GPS-based app, specifically BlindSquare [25]. Participants will be asked to complete navigation tasks (e.g., finding a nearby pharmacy) with both systems. We will compare participants' performance and behaviors in both conditions by collecting system usage logs and conducting semi-structured interviews to understand their overall impressions. We aim to understand the extent to which street cameras can support precise and real-time outdoor navigation for blind pedestrians.

ACKNOWLEDGMENTS

We thank Lindsey Tara Weiskopf for literature review and Arjun Nichani for initial prototyping. We thank our study participants for participating in the study. This work was supported in part by the National Science Foundation (NSF) and Center for Smart Streetscapes (CS3) under NSF Cooperative Agreement No. EEC-2133516, ARO Grant No. W911NF1910379, NSF Grant No. CNS-1827923, NSF Grant No. CNS-2038984, and corresponding support from the Federal Highway Administration (FHWA). Daniel Weiner and Xin Yi Therese Xu were supported by the Columbia–Amazon SURE Program. Sophie Ana Paris, Chloe Tedjo, and Josh Bassin were funded by the NSF Grant No. 2051053 and 2051060.

REFERENCES

[1] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: A navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 90–99. https://doi.org/10.1145/2935334.2935361

[2] Dragan Ahmetovic, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. 2017. Mind Your Crossings: Mining GIS Imagery for Crosswalk Localization. *ACM Transactions on Accessible Computing* 9, 4 (Dec. 2017), 1–25. https://doi.org/10.1145/3046790

[3] Amnesty International. 2021. Surveillance City: NYPD can use more than 15,000 cameras to track people using facial recognition in Manhattan, Bronx and Brooklyn. https://www.amnesty.org/en/latest/news/2021/06/scale-new-york-police-facial-recognition-revealed/

[4] Wen-Chung Chang, Chia-Hung Wu, Wen-Luo Luo, and Huan-Chen Ling. 2013. Mobile robot navigation and control with monocular surveillance cameras. In *2013 CACS International Automatic Control Conference (CACS)*. 192–197. https://doi.org/10.1109/CACS.2013.6734131

[5] Coco Feng. 2019. China the most surveilled nation? The US has the largest number of CCTV cameras per capita. https://www.scmp.com/tech/gear/article/3040974/china-most-surveilled-nation-us-has-largest-number-cctv-cameras-capita

[6] COSMOS Project. 2023. Hardware: Cameras. https://wiki.cosmos-lab.org/wiki/Hardware/Cameras

[7] Ping-Jung Duh, Yu-Cheng Sung, Liang-Yu Fan Chiang, Yung-Ju Chang, and Kuan-Wen Chen. 2020. V-eye: A vision-based navigation system for the visually impaired. *IEEE Transactions on Multimedia* 23 (2020), 1567–1580.

[8] 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)*. Association for Computing Machinery, New York, NY, USA, 323–324. https://doi.org/10.1145/2661334.2661344

[9] John C. Flanagan. 1954. The critical incident technique. *Psychological Bulletin* 51, 4 (1954), 327–358.

[10] Frank Hersey. 2017. China to have 626 million surveillance cameras within 3 years. https://technode.com/2017/11/22/china-to-have-626-million-surveillance-cameras-within-3-years/

[11] Leo A. Goodman. 1961. Snowball Sampling. *The Annals of Mathematical Statistics* 32, 1 (1961), 148–170. https://www.jstor.org/stable/2237615 Publisher: Institute of Mathematical Statistics.

[12] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M. Kitani, and Chieko Asakawa. 2019. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 68–82. https://doi.org/10.1145/3308561.3353771

[13] Richard Guy and Khai Truong. 2012. CrossingGuard: exploring information content in navigation aids for visually impaired pedestrians. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 405–414. https://doi.org/10.1145/2207676.2207733

[14] David L Harkey, Daniel L Carter, Janet M Barlow, and Billie Louise Bentzen. 2007. Accessible pedestrian signals: A guide to best practices. *National Cooperative Highway Research Program, Contractor's Guide for NCHRP Project* (2007).

[15] Microsoft Inc. 2018. Microsoft Soundscape - Microsoft Research. https://www.microsoft.com/en-us/research/product/soundscape/. (2018).

[16] Gaurav Jain, Yuanyang Teng, Dong Heon Cho, Yunhao Xing, Maryam Aziz, and Brian A. Smith. 2023. "I Want to Figure Things Out": Supporting Exploration in Navigation for People with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 63:1–63:28. https://doi.org/10.1145/3579496

[17] 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 (April 2018), 1–28. https://doi.org/10.1145/3178853

[18] Robert K Katzschmann, Brandon Araki, and Daniela Rus. 2018. Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 3 (2018).

[19] Laura Griffin. 2020. Surveillance Cameras Are Everywhere. And They're Only Going To Get More Ubiquitous. https://crimereads.com/surveillance-cameras-are-everywhere-and-theyre-only-going-to-get-more-ubiquitous/

[20] Jaewook Lee, Jaylin Herskovitz, Yi-Hao Peng, and Anhong Guo. 2022. ImageExplorer: Multi-Layered Touch Exploration to Encourage Skepticism Towards Imperfect AI-Generated Image Captions. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–15. https://doi.org/10.1145/3491102.3501966

[21] Xiang Li, Hanzhang Cui, John-Ross Rizzo, Edward Wong, and Yi Fang. 2020. Cross-Safe: A Computer Vision-Based Approach to Make All Intersection-Related Pedestrian Signals Accessible for the Visually Impaired. In *Advances in Computer Vision*, Kohei Arai and Supriya Kapoor (Eds.). Vol. 944. Springer International Publishing, Cham, 132–146. https://doi.org/10.1007/978-3-030-17798-0_13

[22] Yimin Lin, Kai Wang, Wanxin Yi, and Shiguo Lian. 2019. Deep Learning Based Wearable Assistive System for Visually Impaired People. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, Seoul, Korea (South), 2549–2557. https://doi.org/10.1109/ICCVW.2019.00312

[23] Liza Lin, Newley Purnell. 2019. A World With a Billion Cameras Watching You Is Just Around the Corner. https://www.wsj.com/articles/a-billion-surveillance-cameras-forecast-to-be-watching-within-two-years-11575565402

[24] Sergio Mascetti, Dragan Ahmetovic, Andrea Gerino, and Cristian Bernareggi. 2016. ZebraRecognizer: Pedestrian crossing recognition for people with visual impairment or blindness. *Pattern Recognition* 60 (Dec. 2016), 405–419. https://doi.org/10.1016/j.patcog.2016.05.002

[25] MIPsoft. 2016. BlindSquare. https://www.blindsquare.com/

[26] Marko Modsching, Ronny Kramer, and Klaus ten Hagen. 2006. Field trial on GPS Accuracy in a medium size city: The influence of built-up. In *3rd workshop on positioning, navigation and communication*, Vol. 2006. 209–218.

[27] Petr Oščádal, Daniel Huczala, Jan Bém, Václav Krys, and Zdenko Bobovský. 2020. Smart Building Surveillance System as Shared Sensory System for Localization of AGVs. *Applied Sciences* 10, 23 (Nov. 2020), 8452. https://doi.org/10.3390/app10238452

[28] Sabrina A. Paneels, Dylan Varenne, Jeffrey R. Blum, and Jeremy R. Cooperstock. 2013. The Walking Straight Mobile Application: Helping the Visually Impaired Avoid Veering. (July 2013).

[29] Jagannadh Pariti, Vinita Tibdewal, and Tae Oh. 2020. Intelligent Mobility Cane - Lessons Learned from Evaluation of Obstacle Notification System using a Haptic Approach. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–8. https://doi.org/10.1145/3334480.3375217

[30] Roman Pflugfelder and Horst Bischof. 2010. Localization and Trajectory Reconstruction in Surveillance Cameras with Nonoverlapping Views. *IEEE transactions on pattern analysis and machine intelligence* 32 (April 2010), 709–21. https://doi.org/10.1109/TPAMI.2009.56

[31] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2019. WatchOut: Obstacle Sonification for People with Visual Impairment or Blindness. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 402–413. https://doi.org/10.1145/3308561.3353779

[32] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark,

Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. http://arxiv.org/abs/2103.00020 arXiv:2103.00020 [cs].

[33] Lisa Ran, Sumi Helal, and Steve Moore. 2004. Drishti: an integrated indoor/outdoor blind navigation system and service. In *Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the*. IEEE, 23–30.

[34] Dipankar Raychaudhuri, Ivan Seskar, Gil Zussman, Thanasis Korakis, Dan Kilper, Tingjun Chen, Jakub Kolodziejski, Michael Sherman, Zoran Kostic, Xiaoxiong Gu, et al. 2020. Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless. In *Proc. ACM MobiCom*.

[35] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M. Kitani, and Chieko Asakawa. 2019. NavCog3 in the Wild: Large-scale Blind Indoor Navigation Assistant with Semantic Features. *ACM Transactions on Accessible Computing* 12, 3 (Sept. 2019), 1–30. https://doi.org/10.1145/3340319

[36] Jae Shim and Young Cho. 2016. A Mobile Robot Localization via Indoor Fixed Remote Surveillance Cameras. *Sensors* 16, 2 (Feb. 2016), 195. https://doi.org/10.3390/s16020195

[37] Hojun Son, Divya Krishnagiri, V. Swetha Jeganathan, and James Weiland. 2020. Crosswalk Guidance System for the Blind. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. 3327–3330. https://doi.org/10.1109/EMBC44109.2020.9176623 ISSN: 2694-0604.

[38] Wim Taymans, Steve Baker, Andy Wingo, Rondald S Bultje, and Stefan Kost. 2013. Gstreamer application development manual (1.2. 3). *Publicado en la Web* 72 (2013).

[39] Juan Terven and Diana Cordova-Esparza. 2023. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv preprint arXiv:2304.00501* (2023).

[40] Charles Vicek, Patricia McLain, and Michael Murphy. 1993. GPS/dead reckoning for vehicle tracking in the" urban canyon" environment. In *Proceedings of VNIS'93-Vehicle Navigation and Information Systems Conference*. IEEE, 461–34.

[41] Hsueh-Cheng Wang, Robert K. Katzschmann, Santani Teng, Brandon Araki, Laura Giarre, and Daniela Rus. 2017. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Singapore, Singapore, 6533–6540. https://doi.org/10.1109/ICRA.2017.7989772

[42] Shiyun Yang, Emily Bailey, Zhengye Yang, Jonatan Ostrometzky, Gil Zussman, Ivan Seskar, and Zoran Kostic. 2020. COSMOS smart intersection: Edge compute and communications for bird's eye object tracking. In *Proc. SmartEdge*.

[43] Chris Yoon, Ryan Louie, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, and Paul Ruvolo. 2019. Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 210–221. https://doi.org/10.1145/3308561.3353788