Learning from Interactions with Blind Users for Customized and Scalable Navigation Assistance Systems (Final Report)

Eshed Ohn-Bar (PI), Venkatesh Saligrama (Co-PI), Calin Belta (Co-PI),
Ruizhao Zhu (RA), Jimuyang Zhang (RA)

This project funded two PhD students, Ruizhao Zhu and Jimuyang Zhang. The project outcomes have provided the foundation for scalable design and evaluation of intelligent systems that can seamlessly interact with individuals with visual impairments. First, we have been developing a uniquely realistic simulation environment for teaching intelligent systems about the needs individuals with disabilities. Second, we performed an IRB-approved study to guide system and algorithm design. Third, based on our in-simulation and real-world findings, we have been developing a meta-learning framework for *personalization of intelligent systems*, i.e., to enable interaction with diverse real-world settings and end-users. In this process, we have collaborated with a local outreach organization, the Carroll Center for the Blind in Newton, to engage end-users throughout the entire technology development process. Below, we summarize the research conducted as well as discuss our future research and funding plans.

Motivation: How can we design intelligent systems that can seamlessly interact with and assist individuals with visual impairments? Factors related to non-visual reasoning and safety require precise modeling of information needs for determining what, when, and how to best assist an end-user. An elderly blind user, for instance, may benefit from additional guidance and context during challenging navigating tasks, such as those comprising open spaces with frequent obstacles and crowds [4]. Even when providing basic guidance cues, e.g., to turn in a certain direction towards an elevator button or a door handle, a slight delay or premature signaling can result in user confusion and navigation errors. Based on the preliminary findings from our user study, we find Orientation and Mobility guides to carefully tailor the timing and content of their instructions, i.e., to minimize cognitive load and ensure smooth and safe navigation performance. Such real-time adaptation also enables guides to seamlessly accommodate users with various mobility skills (e.g., cane strategies), aids, and personal preferences.

Limitations in Current Systems: There are two fundamental challenges today preventing the development of technologies that can adapt to various needs and preferences of individuals with visual impairments. First, individuals with visual impairments are mostly absent from datasets in computer vision and machine learning. Cost, privacy, and safety issues have led to a lack of standardized development frameworks for developing data-driven adaptive tools, e.g., benchmarks with relevant use-cases and data collection and sharing principles. This limitation hinders even preliminary analysis with respect to current approaches for personalization, e.g., meta-learning [2, 6, 9] Second, factors related to non-visual reasoning and safety requires more elaborate modeling of the information needs of an end-user with visual impairments. This practical limitation has led accessibility researchers to extensively hand-engineer and tune guidance feedback and interaction properties by assistive systems. This current non-scalable process of careful manual design of system guidance properties *hinders system's ability to flexibly accommodate the highly diverse needs and scenarios*. When encountering a new user (e.g., with different mobility skills or aids) or a new environment (e.g., various acoustic and layout properties), the interaction settings must be manually re-adjusted in a cumbersome, non-scalable process [7]. This manual design process is perhaps part of the reason researchers and developers tend to evaluate their carefully engineered prototypical technologies over a relatively small and homogeneous set of users (between 3-10, e.g., [1]), with pre-assumed user and route characteristics and simplified navigation tasks. Indeed, based on our preliminary findings from our on-going project, existing assistive guidance systems are easily confounded when encountering the diverse information needs of naturalistic real-world users and settings. The aforementioned two challenges hinder the development of automatic personalization tools, and result in brittle implementations that fail to anticipate and react to user needs when encountering a new user, device, or scenario. The key technical merit of our project lies in introducing a principled framework and model for facilitating personalized and safe interaction when assisting individuals with visual impairments navigating throughout dense and dynamic settings.

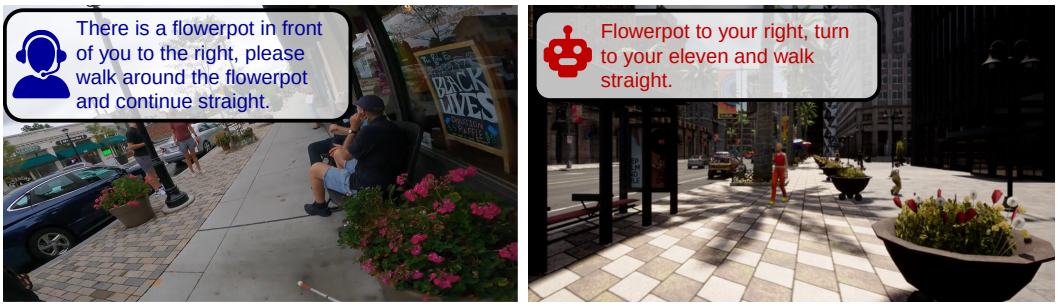

Figure 1. Our goal is to develop intelligent systems that can consider the needs of a walker with visual impairments when seamlessly maintaining situational awareness. Left: Real-world image from the perspective of a participant in our user study, with overlaid navigational instructions provided by an Orientation and Mobility (O&M) expert. Right: First-person view of a simulated pedestrian navigating an urban sidewalk with model generated instructions overlaid.

1. Summary of Completed Research

Despite ample publicly available benchmarks, there are no current datasets suitable for model training and evaluation of timely, safety-critical, and ability-aware navigation guidance to walkers with visual impairments. Towards exploring real-time information needs and fundamental challenges in our novel modeling task, we collect the first multi-modal real-world benchmark with recorded Orientation and Mobility (O&M) experts instructional guidance in dynamic urban walking navigation settings. We then leverage the real-world study to inform the design of a novel realistic simulation environment. Altogether, the two benchmarks will be used to produce complementary analysis while tackling inherent issues in safety, cost, and scalability of real-world data collection with participants with visual impairments. The two benchmarks will also be used for training assistive AI system and comprehensively analyze limitations in current personalization techniques across diverse scenarios (e.g., users, harsh weathers, geographical locations).

Real-World User Study and Dataset: We are the first to collect synchronized multi-modal camera and sensor data together with their corresponding in-situ expert instructional guidance. We recruited 13 participants through the mailing list of a local outreach services center, including 10 individuals with visual impairments and three OM guides (to analyze expert diversity). We sought to collect video and sensor measurements during navigation in real-world urban scenes with expert guidance from the perspective of an assistive system, i.e., a first-person camera. Therefore, in order to capture naturalistic navigation behavior and real-world challenges associated with assistive technologies, we opted for a remote guidance solution. While the limited perspective incorporates a practical challenge, this study design choice also lends to scalability due to minimal mount configuration, ease of data collection, and ultimate large-scale deployments on commodity devices, e.g., smartphones. We asked the participants to navigate an unfamiliar 110m planned route through a busy business district with typical weekday traffic, including pedestrians, vehicles, and shops. We ensured control for confounding factors: participants were called on different days and on varying hours. The equipment included a 5G smartphone, an additional GoPro camera mounted to a chest harness, and a Bluetooth bone-conducting headset to provide instructions without hindering acoustic reasoning. GPS, IMU, audio, and camera data were all captured synchronously. We note that the restricted forward view provided by a chest-mounted camera rarely provided a complete view of the surroundings and potential obstacles. This necessitated crucial collaboration between the navigator and the guide, an interactive functionality that we wish to embody in our assistive agent. For instance, in order to gather sufficient visual information for safe navigation the expert may ask the navigator to stop and scan the environment by rotating their torso to pan and tilt the camera.

Need for Personalization: We empirically found guides to tailor and provide additional feedback, e.g., regarding obstacles and requests to stop, automatically to accommodate certain participants. We also found expert guides to personalize to mobility aids, e.g., with less obstacle descriptors for a participant using a guide-dog compared to those using a walking cane. Previous experience with assistive technologies showed statistically significant differences in the number of interventions along the route as well as navigation time along the walking portion of the route ($p < .05$). The need to tailor instructions across participants and settings was common in participant feedback. For instance, a participant mentioned: *"The best advice I can give is to ask each person how much information they would like, everyone has different preferences for how much information they would like."*. Others expressed explicit preferences, e.g., *"There's a lot my dog can help me go around but it is not the same as somebody telling you where the holes are along the way, the kind of information others find obvious while walking around."*

Simulation Environment for Personalized AI Training: To begin addressing issues in data scarcity in the context of accessibility, we developed an accessibility-centered realistic simulation environment (Fig. 1, right). The simulation supports generation of ample amounts of finely annotated, multi-modal data in a safe, cheap, and privacy-preserving manner with various edge cases, diverse settings, and different walking behaviors. We spawn navigating pedestrians and capture a first-person image perspective together with complete ground-truth information of surrounding landmarks and obstacles (i.e., 3D location of buildings, pedestrians, sidewalks, trees, etc.). Given a current walker position, a sampled goal, and a constructed Bird's-Eye-View (BEV) image, we extract walkable space and obtain a path using A* planning. We then employ the planned path to construct instructional sentences. We contextualize the instructions by extracting surrounding obstacle information from the BEV along the path and inform regarding obstacles in proximity (e.g., pedestrians, building). While this process can be used to generate standardized instructions, we leverage insights from our real-world study together with prior literature in orientation and mobility strategies to consider relevant navigation strategies and immediate information needs. For instance, we leverage clock orientation to indicate turning which has been found to be more intuitive for users with visual impairments.

Meta Imitation Learning for Personalization: Manually designing a generalized assistive instruction generation expert is challenging. To learn to generate human-like instructions, we assume a dataset with expert instructional guidance (available in our real-world and synthetic benchmarks). We can then optimize a model to imitate expert instructions, i.e., using supervised learning [5, 8]. We then leverage meta-learning [2, 3, 6, 9] to personalize the model. Meta learning is inspired by humans' learning ability of adapting their knowledge of representations, beliefs and predictions as they encounter new tasks. This is typically accomplished by training a meta-model on a diverse set of tasks, such that the meta-model can in turn train and output a model on a new task using only a few training examples. In our context, system usability pivots on the system's ability to quickly personalize its model from little data. Our preliminary results of personalization over the real-world data suggest significant performance gains. Within five samples, the model's accuracy when imitating the expert increases by about 200% (on average across all participants). However, despite the improvement, we find absolute model performance to be quite low for our assistive instruction task (5-15% accuracy with across various language metrics and participants). Moreover, personalization can also worsen performance for a subset of the participants. As our work provides the first step towards automatically personalized assistive navigation systems, future work can now use our tools to rigorously analyze and tackle such issues. We are also organizing a workshop this year to begin tackling fundamental issues in model design and evaluation for assisting individuals with visual impairments.

2. Future Plans

We plan to continue working on the meta learning framework in our simulation so it may be validated safely in a follow-up real-world study. We will be working on this during the Summer and Fall semesters. We will then submit a paper with our novel benchmarks and findings, targeting machine learning and computer vision conferences, such as the International Conference on Learning Representations and the International Conference on Computer Vision and Pattern Recognition. Various aspects of this project, from the simulation environment and up to meta learning for personalization, will also become part of Jimuyang's and Ruizhao's PhD dissertations. We will target relevant funding in the Fall, such as the NSF programs for Smart and Connected Communities (SCC) and Computer and Information Science and Engineering (CISE).

References

[1] N. Fallah, I. Apostolopoulos, K. Bekris, and E. Folmer. Indoor human navigation systems: A survey. *Interacting with Computers*, 25(1):21, 2013. 1

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017. 1, 3

[3] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40, 2017. 3

[4] E. Ohn-Bar, J. Guerreiro, K. Kitani, and C. Asakawa. Variability in reactions to instructional guidance during smartphone-based assisted navigation of blind users. In *International Joint Conference on Pervasive and Ubiquitous Computing*, 2018. 1

[5] Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, and Jan Peters. An algorithmic perspective on imitation learning. *arXiv preprint arXiv:1811.06711*, 2018. 3

[6] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *ICML*, 2017. 1, 3

[7] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa. NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In *ASSETS*, 2017. 1

[8] Wen Sun, Arun Venkatraman, Geoffrey J Gordon, Byron Boots, and J Andrew Bagnell. Deeply aggrevated: Differentiable imitation learning for sequential prediction. In *ICML*, 2017. 3

[9] Pan Zhou, Xiaotong Yuan, Huan Xu, Shuicheng Yan, and Jiashi Feng. Efficient meta learning via minibatch proximal update. In *NeurIPS*, 2019. 1, 3