Vermessung & Geoinformation 3/2018, S. 220 – 225, 1 Abb.

Pedestrian Navigation: What Can We Learn from Eye Tracking, Mixed Reality and Machine Learning

Was können wir aus Eye Tracking, Mixed Reality und Machine Learning lernen?

Ioannis Giannopoulos, Wien

Abstract

Understanding how humans interact with their surroundings during spatial decision-making is crucial for the understanding of several processes, such as navigation. Furthermore, during spatial decision-making, humans also interact with spatial data often presented to them through a display device. Through eye tracking, mixed reality and machine learning we are able to come closer to an understanding, optimize the relevant interaction dialogues, classify relevant interaction spaces and assist humans during the process of spatial decision-making.

Keywords: Navigation, Eye Tracking, Mixed Reality, Machine Learning

Kurzfassung

Um verschiedene Prozesse wie zum Beispiel die Navigation zu verstehen, ist es entscheidend zu verstehen wie Menschen mit ihrer Umgebung während der Entscheidungsfindung interagieren. Während der räumlichen Entscheidungsfindung interagieren Menschen auch mit räumlichen Daten, die Ihnen oft über Display Geräte präsentiert werden. Mithilfe von Eye Tracking, Mixed Reality und Machine Learning sind wir in der Lage, ein besseres Verständnis und eine Optimierung der relevanten Interaktionsdialogen zu erzielen, relevante Interaktionsräume zu klassifizieren sowie Menschen während des Entscheidungsfindungsprozesses zu assistieren.

Schlüsselwörter: Navigation, Eye Tracking, Vermischte Realität, Maschinelles Lernen

1. Introduction

Human navigation is defined by Golledge [1] as the goal directed movement in space, in order to reach a predefined destination. According to Montello [2], navigation consists of two components, namely locomotion and wayfinding. During locomotion we make decisions in our direct surroundings, e.g., trying to avoid obstacles. During wayfinding we make a series of decisions, also incorporating survey knowledge, which eventually leads us to the desired destination.

Human navigation is an interdisciplinary research topic in the focus of several domains, ranging from Cognitive Psychology to Computer Science and Geoinformation. Human navigation provides a content rich testbed allowing answering many different kinds of research questions which can also easily be generalized to further domains. Researchers investigate the factors that lead to successful navigation in terms of efficiency, effectiveness, user experience, cognitive load, amongst others. These factors are not only user dependent, e.g., spatial abilities, but also environment related, e.g., decision points.

The research experiments in this domain are performed in the real environment, but also in lab settings. One of the advantages of experiments in the lab is the control over the experiment, being able to observe and understand where the effects are coming from. Although often claimed that there is a lack in external validity, studies have shown that lab experiments externalize quite well [3]. Furthermore, the advantages that a lab setting can offer are tremendous. Due to the technological advancements of the last decade, a lab can be equipped with several emerging technologies, e.g., eye tracking, body tracking, holographic-like interaction glasses, etc., allowing to observe and analyze the user behavior and interaction with space in more depth. Although some of these technologies can also be utilized for experiments in outdoor environments, there are trade-offs, e.g., sunlight interference is a problem when performing experiments using a mobile eye tracking device [4]

Many researchers have split the relevant research parts in order to make them more manageable and have also proposed classifications and models in order to provide formalisms. Jan Wiener et al. [5] proposed in their work a taxonomy of wayfinding tasks, classifying wayfinding in aided and unaided. Kiefer et al. [6] proposed a wayfinding grammar as a mean to cover and describe the wayfinding processes in a formal yet modular way. Giannopoulos et al. [7] characterized the relevant space by introducing three components,

1) „the space the user interacts in (i.e., the position of the user)",
2) *„the spatial information the user interacts with (e.g., the information on the map)"* and
3) *„the space the user interacts with (i.e., the objects in the environment)"*. In the present work we focus on aided wayfinding as well as the spatial information and the space the user interacts with.

In the following sections a brief overview will be given concerning the factors that constitute to the complexity of wayfinding, the current research directions and how open issues are currently addressed. Furthermore, eye tracking, mixed reality and machine learning will be discussed in order to emphasize the benefits that can arise through them when doing research in navigation.

2. Wayfinding Complexity

While navigating in unfamiliar environments, we often make use of assistance aids, either of traditional cartographic paper maps or of emerging mobile technologies, visualizing the instructions on a screen. During wayfinding we try to interpret and match these instructions to the surrounding environment, e.g., while approaching a decision point. This process already highlights three factors that are immediately essential for the success of this matching task, namely, the complexity of the surrounding environment, the complexity of the given instruction and of course, the abilities of the user that has to perform these actions. Giannopoulos et al. [8] call this a *Wayfinding Decision Situation*: *„A wayfinding decision situation occurs when a specific wayfinder has to make a wayfinding decision in a certain environment with a certain instruction."* In their work, they provided the first model to approximate the complexity of a decision situation (see Equation 1) by integrating the complexity of the environments in terms of branches that are possible to be chosen at a decision point, the complexity of the instructions at hand, as well as based on the spatial abilities of the wayfinder. The wayfinder has to interpret the instructions, match them into the environment and finally, make a decision.

$$c(e, t_i, U) = c(e) \oplus c(t_i) \oplus f(U) \Rightarrow$$
$$c(e, t_i, U) = w_1 * \#br + w_2 * (\beta * adv_{vis} + (1-\beta) * lm) + w_3 * sa$$

sa: spatial abilities, br: number of branches, adv_{vis}: advance visibility and lm: landmark matching

$$c(e, t_i, U) \in [0, 1]$$

Equation 1: *The Wayfinding Decision Situation Complexity model defined by Giannopoulos et al. [8]*

Each of the three factors integrated in the above model has an impact on the wayfinding complexity. It is important to understand that the easier the interaction between the user and the surrounding environment and the instructions at hand, the lower the complexity [8]. According to Stephan Winter [9] and his introduced measure for advance visibility, the wayfinding complexity increases the longer it takes to see the landmark used in an instruction. Furthermore, Giannopoulos et al. [10] demonstrated that the timing of the instructions also has an impact on the complexity. According to their research, the optimal timing of instructions is also strongly dependent on the wayfinding situation. In their work they introduced a formal model based on a survival analysis, modeling the optimal timing using an *Accelerated Failure Time* (AFT) model. By using the estimated parameters of the model and filling in the input parameters, e.g., age of the user, type of decision point, etc., a probability distribution is computed which can be utilized to assess the optimal timing of an instruction for a specific wayfinding situation.

Current research tries to optimize human navigation by optimizing the complexity of a wayfinding decision situation, either focusing on understanding the strategies wayfinders apply in order to make a decision, or by optimizing the interaction dialogue between the wayfinder, the aid and the environment, amongst other approaches.

3. Human Computer Interaction in Navigation

One branch of current research focuses on optimizing the interaction with the assistance aid in order to outsource the complexity to the processing unit of the device. Several approaches have been evaluated through empirical experiments focusing mostly on optimizing the navigation time, i.e., the time to reach the desired destination from the origin, as well as spatial knowledge acquisition, i.e., learning the environment. Next to this, there is also a focus on user experience and in reducing the cognitive load during decision making. For instance, Giannopoulos et al. [11] introduced the *GazeNav* concept, a gaze-based pedestrian navigation assistance system, focusing on reducing the navigation errors by minimizing the necessary interaction with the environment. Another approach was introduced by Gkonos et al. [12] and Pielot et al. [13] focusing on vibrotactile interaction, providing the navigation instructions through vibrations on a custom developed waist belt. Both approaches, the gaze-based and vibrotactile approach are based on non-visual navigation, i.e., they do not require that the user reads the instructions from a screen. There are of course also several visual-based approaches which on their part focus on optimizing the interaction with the mobile screen, e.g., to optimally obtain and understand the provided instructions. Giannopoulos et al. [14] introduced in their work the *GeoGazemarks* concept, a gaze-based interaction that allows to keep marks of the orientation on mobile maps. Another very successful approach that is followed in order to provide navigation assistance focuses on optimizing the interaction with the surrounding environment. Anagnostopoulos et al. [15] introduced in their work the gaze-informed Location Based Services. The user is able to obtain information by just gazing at elements of the environment. For instance, while looking at a specific building the user could obtain information that could help them reach the desired destination.

4. Eye Tracking in Navigation

Navigation requires that the wayfinder interacts with the surrounding environment and collects information that will be included into the decision-making process. While humans acquire this information through multiple senses, the visual sense is of great importance [16]. As addressed by Kiefer et al. [16], the main reasons why eye tracking can be very beneficiary for navigation research are the following:

1) *the majority of today's wayfinding aids are based on visual interaction*
2) *visual attention can easily be measured (which is not the case for other senses)*
3) *the information acquisition process can be measured*

Eye tracking technology exists already for several decades, with research in the 60s' already focusing on what our eye movements can reveal. For instance, Yarbus [17] in his research showed that eye movements are task dependent, a finding that inspired recent research to focus on activity recognition based solely on eye movements. Kiefer et al. [18] captured eye movements of participants during map reading tasks and used them in order to classify and automatically detect the map reading tasks while they occur.

Eye tracking is no longer only remote, but also mobile. The typical eye tracking technology, i.e., remote eye tracking, is mostly fixed under a desktop monitor, allowing to perform experiments in a lab environment in front of a computer screen. Mobile eye tracking technology allows to perform experiments in the wild, allowing the participant to freely move while the gaze is recorded. Remote eye trackers have typically a higher precision, accuracy and recording frequency than mobile eye trackers, allowing to analyze eye movements more thoroughly. From the captured eye movements, the eye events fixations and saccades are computed and typically used for further analysis. Eye fixations occur when our eyes remain relatively still, which is the case when we obtain information. The rapid eye movements in between are the saccades.

Eye tracking technology is commonly used during navigation experiments in order to understand the process of wayfinding. According to Downs and Stea [19] the process of wayfinding can be split into four sequential and interrelated steps, namely, *orientation*, *route choice*, *route monitoring* and *recognition of the destination*. Kiefer et al. [4] utilized mobile eye tracking technology in order to investigate the process of self-localization. Through a user experiment, they were able to observe the strategies that were applied during self-localization and further distinguish some of the reasons that lead to incorrect localization. Unsuccessful participants were mostly focusing on elements on the map which were not visible from the experiment area. The proper choice of landmarks in a navigation instruction can therefore be of immense importance and is subject to research where eye tracking technology can provide helpful insights in understanding what is considered a landmark and which landmarks are appropriate for visual representation in assistant aids. For instance, Viaene et al. [20] focused on identifying the objects in the environment that can be categorized as landmarks. In their experiment they tracked the wayfinders gaze and proposed a gaze-based measure in order to classify objects as landmarks. Franke et al. [21] performed an eye tracking user study in order to identify the most effective ways to represent landmarks on maps and Ohm et al. [22] performed an eye tracking study in order to evaluate landmark-based interfaces for indoor navigation.

Navigation research is not limited to the outdoor urban environments, but focuses on indoor environments as well. Schnitzler et al. [23] utilized mobile eye tracking glasses and performed a user experiment in a large complex indoor environment in order to investigate how wayfinders use different navigation assistance aids in complex multi-level environments and Schrom-Feiertag et al. [24] performed an eye tracking evaluation of an indoor guidance system.

This brief overview of eye tracking research in navigation demonstrates some of the benefits that can arise when utilizing eye tracking. Eye tracking technology can be used to gain an understanding of the wayfinding processes and the user strategies in order to make decisions. Furthermore, eye tracking can also be utilized as an interaction modality, enabling implicit and explicit interaction [25] with the environment and the assistance aid in order to optimize navigation.

5. Mixed Reality in Navigation

Due to recent technological advances in Mixed Reality, the functionalities and possible applications that can evolve become interesting for navigation research. Mixed Reality (MR) is constituted by the combination of two technologies, Augmented Reality (AR) and Virtual Reality (VR). MR can be found in between these two worlds. According to Azuma [26] an MR application has to „combine real and virtual objects, to provide interactivity in real time and be registered in 3D”. Head-mounted displays are used for the implementation of the MR experience, allowing for a more realistic and immersive perception and interaction with the virtual objects [27] (see Figure 1). MR is already utilized for research in several research fields, ranging from hologram-like visualizations in MR for creative self-expression [28] to remote collaboration using virtual objects as spatial cues [29], [30].

For research in Navigation, MR gets specifically interesting for several reasons. During aided navigation, the information provided by the aid is displayed in two dimensions, although very often the information has three dimensions. This requires from the wayfinder to perform a mental rotation and understand the projected information in order to make sense of the third dimension, unnecessarily increasing the wayfinding complexity even more. In the MR domain, this information could be projected in the real environment in front of the user, easing the incorporation of this information in the decision-making process. In a similar manner, instructions, e.g., landmarks could directly be placed in the real environment, removing the necessity of relying on real environment landmarks.

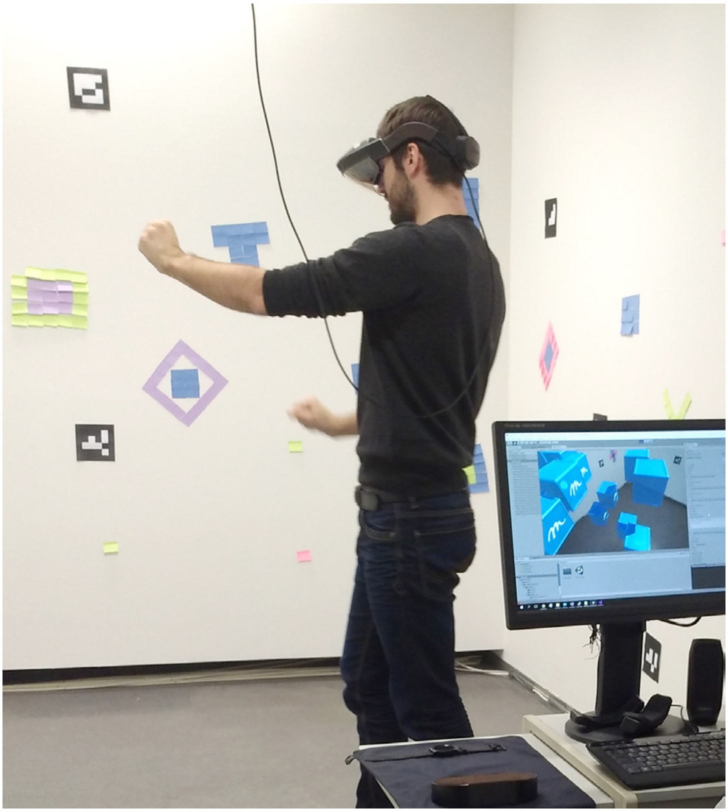

Fig. 1: *A user interacting with spatial objects in a Mixed Reality Environment. Cubes are projected in front of the user providing a holographic-like experience, allowing to move around the projection.*

Furthermore, MR allows to investigate in depth how humans interact with space, allowing to evaluate multiple instances of spatial elements easily, by simply changing the 3D visualization. For instance, MR can allow to verify theories concerning elements of an urban environment that can be useful in order to ease the wayfinding process. These elements do not have to be already incorporated in the environment, they can simply be projected, making the experimental process faster, cheaper and even possible, not having to rely on the existing urban infrastructure.

6. Machine Learning in Navigation

Machine learning is defined as a computer program that can *„learn from experience E with respect to some class of Tasks T and performance measure P, if its performance and tasks in T, as measured by P, improves with experience E"* [31, p. 2]. In navigation research, where a lot of the tasks are still difficult to clearly distinguish while they occur, machine learning can be very helpful. For instance, lets assume that the set of relevant Tasks T are the ones defined in the Introduction, namely orientation, route choice, route monitoring and recognition of destination. It would be very beneficiary if we could track down the sequence and frequency of these interrelated steps. For instance, we could look at the results of the algorithm and observe when and how often these steps occurred and perform even more targeted analyses, being able to interpret the experimental observations with respect to the user activity (e.g., the user was gazing at these specific elements during orientation). Liao et al. [32] applied machine learning in their work using eye tracking data in order to classify typical navigation tasks in the real environment. Kiefer et al. [18] applied machine learning on eye movement data of users interacting with a map in order to recognize user activities on cartographic maps, such as route planning. The results look very promising, supporting even more the above case.

Machine learning can be useful for several stages of the wayfinding process. For instance, identifying the familiarity level of the user with the environment can be very beneficiary for an assistance system which in turn can adapt according to this information. Machine learning and eye tracking can also be utilized to detect the cognitive load of wayfinders. Duchowski et al. [33] termed the Index of Pupillary Activity, is shown to discriminate task difficulty vis-à-vis cognitive load (if the implied causality can be assumed utilized in their work eye pupil data in order to detect the user´s cognitive load level.

7. Conclusions

Past and current research in navigation has already successfully demonstrated how technologies such as eye tracking can help understanding the processes involved. This work tried to highlight the benefits that can arise from this and other technologies in navigation research and provide examples where eye tracking, mixed reality and machine learning could help to get deeper insights or even make it possible to perform research in environments that do not yet exist in reality.

References

[1] R. G. Golledge, "Human wayfinding and cognitive maps," in *Wayfinding behavior*: *Cognitive mapping and other spatial processes*, R. G. Golledge, Ed. Baltimore, MA: Johns Hopkins University Press, 1999, pp. 5–45.

[2] D. R. Montello, "Navigation.," in *Cambridge handbook of visuospatial thinking.*, 2005, pp. 257–294.

[3] C. A. Anderson, J. J. Lindsay, and B. J. Bushman, "Research in the Psychological Laboratory: Truth or Triviality?," *Curr. Dir. Psychol. Sci.*, vol. 8, no. 1, pp. 3–9, 1999.

[4] P. Kiefer, I. Giannopoulos, and M. Raubal, "Where Am I? Investigating Map Matching During Self-Localization With Mobile Eye Tracking in an Urban Environment," *Trans. GIS*, vol. 18, no. 5, pp. 660–686, 2013.

[5] J. M. Wiener, S. J. Büchner, and C. Hölscher, "Taxonomy of human wayfinding tasks: A knowledge-based approach," *Spat. Cogn. Comput.*, vol. 9, no. 2, pp. 152–165, 2009.

[6] P. Kiefer, S. Scheider, I. Giannopoulos, and P. Weiser, *A wayfinding grammar based on reference system transformations*, vol. 9368. 2015.

[7] I. Giannopoulos, P. Kiefer, and M. Raubal, "Mobile Outdoor Gaze-Based GeoHCI," in *GeoHCI Workshop at CHI 2013, 2013*, pp. 1--2.

[8] I. Giannopoulos, P. Kiefer, M. Raubal, K. Richter, and T. Thrash, "Wayfinding Decision Situations : A Conceptual Model and Evaluation," in *Geographic Information Science*, 2014, pp. 221–234.

[9] S. Winter, "Route Adaptive Selection of Salient Features," in *Spatial Information Theory. Foundations of Geographic Information Science*, 2003, vol. 2825, pp. 349–361.

[10] I. Giannopoulos, D. Jonietz, M. Raubal, G. Sarlas, and L. Stähli, *Timing of pedestrian navigation instructions*, vol. 86. 2017.

[11] I. Giannopoulos, P. Kiefer, and M. Raubal, "GazeNav: Gaze-Based Pedestrian Navigation," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices & Services*, 2015, pp. 337–346.

[12] C. Gkonos, I. Giannopoulos, and M. Raubal, "Maps, vibration or gaze? Comparison of novel navigation assistance in indoor and outdoor environments," *J. Locat. Based Serv.*, vol. 11, no. 1, 2017.

[13] M. Pielot, B. Poppinga, W. Heuten, and S. Boll, "PocketNavigator : Studying Tactile Navigation Systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 3131–3139.

[14] I. Giannopoulos, P. Kiefer, and M. Raubal, "GeoGazemarks: Providing gaze history for the orientation on small display maps," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 165–172.

[15] V. Anagnostopoulos, M. Havlena, P. Kiefer, I. Giannopoulos, K. Schindler, and M. Raubal, "Gaze-Informed location-based services," *Int. J. Geogr. Inf. Sci.*, vol. 31, no. 9, 2017.

[16] P. Kiefer, I. Giannopoulos, M. Raubal, and A. Duchowski, "Eye tracking for spatial research: Cognition, computation, challenges," *Spat. Cogn. Comput.*, vol. 17, no. 1–2, 2017.

[17] A. L. Yarbus, "Eye movements during perception of complex objects," in *Eye movements and vision*, S. US, Ed. Springer, 1967, pp. 171–211.

[18] P. Kiefer, I. Giannopoulos, and M. Raubal, "Using eye movements to recognize activities on cartographic maps," in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, 2013, pp. 478–481.

[19] M. R. Downs and D. Stea, "The World In The Head," in *Maps in Minds: Reflections on Cognitive Mapping*, Harper & Row, 1977, pp. 99–145.

[20] P. Viaene, P. Vansteenkiste, M. Lenoir, A. De Wulf, and P. De Maeyer, "Examining the validity of the total dwell time of eye fixations to identify landmarks in a building," *J. Eye Mov. Res.*, 2016.

[21] C. Franke and J. Schweikart, "Mental representation of landmarks on maps: Investigating cartographic visualization methods with eye tracking technology," *Spat. Cogn. Comput.*, 2017.

[22] C. Ohm, M. Müller, and B. Ludwig, "Evaluating indoor pedestrian navigation interfaces using mobile eye tracking," *Spat. Cogn. Comput.*, 2017.

[23] V. Schnitzler, I. Giannopoulos, C. Hölscher, and I. Barisic, "The Interplay of Pedestrian Navigation, Wayfinding Devices, and Environmental Features in Indoor Settings," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, 2016, pp. 85–93.

[24] H. Schrom-Feiertag, V. Settgast, and S. Seer, "Evaluation of indoor guidance systems using eye tracking in an immersive virtual environment," *Spat. Cogn. Comput.*, 2017.

[25] A. Schmidt, "Implicit human computer interaction through context," *Pers. Technol.*, vol. 4, no. 2–3, pp. 191–199, 2000.

[26] R. T. Azuma, "A survey of augmented reality," *Presence Teleoperators Virtual Environ.*, 1997.

[27] L.-H. Lee and P. Hui, "Interaction Methods for Smart Glasses," *ACM Comput. Surv.*, 2017.

[28] J. Amores and J. Lanier, "HoloARt: Painting with Holograms in Mixed Reality," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 421–424.

[29] J. Müller, R. Rädle, and H. Reiterer, "Virtual Objects As Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load," *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst.*, 2016.

[30] J. Müller, R. Rädle, and H. Reiterer, "Remote Collaboration With Mixed Reality Displays: How Shared Virtual Landmarks Facilitate Spatial Referencing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 2017.

[31] T. M. Mitchell, *Machine Learning*. 1997.

[32] H. Liao, W. Dong, H. Huang, G. Gartner, and H. Liu, "Inferring user tasks in pedestrian navigation from eye movement data in real-world environments," *Int. J. Geogr. Inf. Sci.*, 2018.

[33] A. T. Duchowski, K. Krejtz, I. Krejtz, C. Biele, A. Niedzielska, P. Kiefer, M. Raubal, and I. Giannopoulos, "The Index of Pupillary Activity: Measuring Cognitive Load vis-à-vis Task Difficulty with Pupil Oscillation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 2018.

Contact

Prof. Dr. Ioannis Giannopoulos, Vienna University of Technology, Department of Geodesy and Geoinformation, Research Group Geoinformation, Gußhausstraße 27-29, A-1040, Vienna.

Email: igiannopoulos@geo.tuwien.ac.at