Analyzing Dynamic Outdoor Eye-Tracking Data

Brent Chamberlain[1], David Evans[2], Laura Schalbetter[3], Peter Kiefer[4],
Adrienne Grêt-Regamey[3], Ulrike Wissen Hayek[3]

[1]Utah State University, Landscape Architecture & Environmental Planning (LAEP), Utah/USA · brent.chamberlain@usu.edu
[2]Utah State University, Landscape Architecture & Environmental Planning (LAEP), Utah/USA
[3]ETH Zurich, Planning of Landscape and Urban Systems (PLUS), 8093 Zurich/Switzerland
[4]ETH Zurich, Geoinformation Engineering, 8093 Zurich/Switzerland

Abstract: Dynamic eye tracking in 3D spaces provides a promising method for understanding human perception and judgment of landscape qualities. Studies applying this technology to outdoor landscapes remain scarce. This experimental study combines outdoor eye-tracking technology with artificial intelligence (computer vision) to represent, visualize, and interpret human observational patterns in a rehabilitated agricultural landscape. It examines what individuals observe during free-viewing of an outdoor environment and how these observations relate to landscape elements. Findings reveal consistent patterns (interest in mid-ground, water features) that align with established landscape theories and highlights phenomena requiring further investigation to enhance the understanding of the relationship between design theories and observational behavior.

Keywords: Outdoor eye tracking experiment, in-situ study, rural environment, gaze patterns, artificial intelligence, computer vision

Subject area: Algorithmic Design and Analysis of Landscapes

1 Introduction

Environmental elements and individual traits shape spatial understanding (SIEGEL & WHITE 1975). Recent research has improved our understanding of how individuals perceive space, considering tasks, capabilities, and environmental elements (STITES et al. 2020, BRUNS & CHAMBERLAIN 2019, LIU & NIJHUIS 2020). This understanding can help steer landscape and urban development towards livable, pleasant and identity-creating places that can foster human well-being (SWITALSKI & GRÊT-REGAMEY 2021). Dynamic eye tracking in 3D spaces is a promising approach to foster the understanding of how environmental elements correlate with human perception and judgements of landscapes (CHE et al. 2023). However, dynamic eye tracking presents challenges in data analysis and interpretation (COTTET et al. 2018). Eye tracking has been used in a wide variety of outdoor spaces for architecture and urban design (AALTO & STEINERT 2024), but there are limited studies related to more naturalized landscapes.

We designed an empirical study with an in-situ dynamic exploration of a rural landscape using eye tracking to assess the relationship between observational patterns and landscape form. In this work, “dynamic” refers to the freedom to observe in any direction from a fixed location. The study aimed to address challenges arising from limitations in eye tracking hardware and software to concisely aggregate and visualize points of fixation in large 3D landscapes. The research questions guiding this work are: 1) how can we analyze eye tracking data of large 3D landscapes, and 2) what are people observing in an outdoor rural environment? The results can advance the analysis of human perception and responses to landscape environments.

Journal of Digital Landscape Architecture, 10-2025, pp. 443-450. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-754-0, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537754042.

2 Methods

2.1 Study Site and Procedure

The study site was selected because it offers several distinct elements and features, including paths, fenced fields, agricultural buildings, and a managed landscape containing areas of native plants, small ponds, and agricultural zones. Further, the area has undulating topography and is surrounded by a woodland with several individual trees spread throughout the interior fields. The view is surrounded by a woodland, which offers a controlled spatial extent to limit potential distractions more common in urban areas.

Participant recruitment was conducted through convenience sampling via a wide email distribution, professional and personal connections directly, as well as indirectly through forwarding and word-of-mouth. Participants had no vision impairments (or were wearing corrective contact lenses). Most participants had completed a university degree and nearly half had a master's degree. While we did not explicitly collect professional backgrounds, we know participants had a variety of academic backgrounds, including spatial planning, architecture, and engineering. Prior experience with the site was not a restricting factor, but we did collect data about prior experience with the site. All procedures were approved through the ETH Zurich Ethics Commission on June 17, 2024 (EK 2024-N-190).

All participants completed the same procedures, and made the observations from the same location. The time of day differed (9:00-16:00) based on appointment and availability. Experimental procedures were completed off-site (about a 5-7minute walk away) from where participants were not able to see any features of the study site. Participants were equipped with the eye-tracker Tobii Pro Glasses 3 (https://www.tobii.com). The glasses track eye movements for both eyes, with a manufacturer-measured 0.6 degrees of error and a sampling rate of 100 Hz. The glasses include an embedded video camera. To control the acoustic environment, participants were fitted with noise-cancelling earbuds that were playing white noise. Participants were then escorted to a small observation platform at a small pond in the middle of the site. They were escorted by walking alongside a moderator who remained at the site. Before the walk, their vision was constrained so they could only see about a meter or two in front of their feet. Once there, they were instructed to freely look in 360° for 5 minutes to visually explore the landscape. The moderator stayed behind the participant during the entire experiment to not distract their view.

2.2 Data Analysis

Data used in this project came from three sources: 1) the Tobii Glasses 3 Eye Tracking hardware, 2) real-time video produced from the glasses, and 3) imagery used to generate panoramas. For the first source, data include two 3D vectors (one for each eye). The vector data are mapped to a coordinate system relative to the second source, which is a video recorded from the center of the glasses. Typically, these two data sources are then used in conjunction with the Tobii Pro Lab to analyze the data. The software Tobii Pro Lab (https://connect.tobii.com/s/lab-downloads) offers assisted mapping functions to align the gaze data to single frames of the videos recorded simultaneously during the eye tracking.

However, this approach can be very labor intensive, especially when the participant is moving in space. Tobii's software requires the researcher to match periods of time from the eye tracker video to a reference image (usually of a small resolution, such as 1920 × 1080 pixels).

Eye-tracking studies are typically done in more controlled settings where the observational spaces can be constrained, making matching easier. For this study, users had a full panoramic experience and exercised a lot of visual freedom. So, matching video segments to a single image would require tedious review of every video segment that matched to a single image that would later make up the entire panorama.

For this study, we propose a method for analyzing outdoor mobile eye-tracking data. Overall, the process includes: collecting site imagery, stitching images into a panorama, collecting eye-tracking data, transposing these data onto the panorama image, cleaning the data, and producing analyses. The first major hurdle was producing the panorama image from photos captured on site. Twenty-four photos (6960 × 4640 pixels) were taken from the participants' standpoint on the platform at eye height (1.5m) with a Canon EOS 90D on a tripod. The panorama was developed using PTGui. This software minimizes horizontal distortions (important for transposing coordinates). The final image covered 360° horizontally and approximately 56° vertically (7538 × 1162 pixels) as shown in Figure 1.

With the panorama created, we needed to transpose eye-tracking coordinates for every frame from the Tobii glasses video to the panorama. Fig. 1 illustrates this concept by matching a single frame (1920 × 1080 pixels) with the original coordinate of the convergence vector from the eye tracker (in red). In this example, the convergence point would have been located at (1800, 900) on the frame, but when transposed to the panorama, has the coordinates (3500, 1000). To accomplish this, we used OpenCV's AKAZE method (SHARMA et al., 2023). OpenCV provides a library of programming functions for real-time computer vision. The AKAZE method uses key points found in both images, then builds a projection to correctly produce the transposed coordinate. AKAZE automatically identified numerous key points within each image that are statistically unique. Then it matches the key points from one image to the other. The distribution of key points and their spatial relationship provides a projection distortion that allows the X,Y pair in the original image to be placed as an X,Y pair in the panorama image.

Fig. 1: Panorama image of the site illustrating the total dimension of the image. The white box represents a single frame from the video that matches the location on the panorama image. The red dot is the convergence point of one observation within that frame. The transposed coordinate is identified in red.

After all data were transposed, they needed to be cleaned so we could identify points of fixation. A fixation allows the eyes to take in detailed information that can then be visually processed (we also included data considered a smooth pursuit as this accounts for head movements). In the cleaning process, raw gaze data were resampled using a linear interpolation method to exactly 100 hz. Blinks were removed (eye openness < 40%). Data gaps longer than 500ms were flagged and not interpolated to minimize potential bias. Data were then clustered using a velocity-based threshold for fixation identification. This method measures the angular velocity of movement from one eye tracking point to another. For this analysis, we excluded all data where the angular velocity exceeded 100 degrees per second in any direction, to isolate only the data where visual processing was likely to be occurring (SALVUCCI & GOLDBERG 2000). Angular velocity is defined here as the overall movement in degrees per second combining both horizontal and vertical components.

The resulting fixation data provide an opportunity to measure what elements individuals view. To analyze these data, we segmented the panorama image using semantic descriptions (see Fig. 2). We categorized the water regions (pond) distinctly because these have shown to garner interest (LIU et al. 2021). Infrastructure has also been segmented (structures & dock), while the treeline and fields make up the middle ground (APPLETON & LOVETT 2003, LIU & NIJHUIS 2020) and the sky as the remaining element.

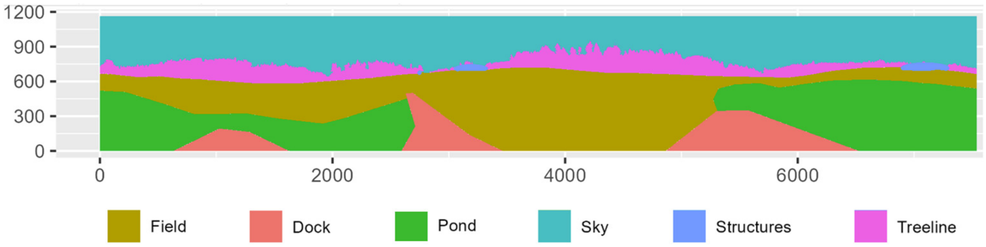

Fig. 2: Semantic segmentation of features in the panorama. Axes are pixel counts.

3 Results

Twenty-six healthy-feeling participants (gender balanced) aged between 22 – 74 (mean age about 36), with normal or corrected-to-normal vision with contact lenses (glasses can obstruct the eye-tracking devices), and normal or corrected-to-normal hearing participated in the study conducted from 20 August 2024 – 2 September 2024. Nearly all participants (23 of 26) had no experience at the site. The transposed data from the eye tracker to the panorama included over 600,000 raw gaze measurements, which yielded 350,000 fixations to be included in the analysis. Figure 3 represents fixations using a panorama of the scene overlaid with a transparency-based mask to convey density (DUPONT et al. 2015). This image was created by developing a heatmap of fixation points (typically visualized as red for high fixation and white as no fixation) but then replacing color with the amount of transparency. The underlying heatmap is a kernel density estimation with an axis-aligned bivariate normal kernel and a bandwidth of 1% along both axes (VENABLES & RIPLEY 2002). More simply, the transparency mask shows the chances that a random participant had a fixation on a location in the scene, as measured by the density of fixations per cell when dividing the image into a 100 × 100 cell grid.

Fig. 3: A transparent gaussian heatmap. The clearer the image, the greater time participants spent fixating at something in the landscape.

The semantic assessment of these data offers another way to represent which kinds of features individuals observed more often than others. Figure 4 shows the ratios of the percent of fixation points (i. e. a proxy for total dwell time given a constant sampling rate) in each semantic category identified in the segmented image (Fig. 1), relative to the proportion of the panorama covered by that semantic category. Given a strictly random distribution of fixations, the ratios would tend towards 1, but these ratios suggest that the field and pond received a disproportionate amount of attention for the degree to which they occupied the field of view. Statistical tests (Levene; Kruskal-Wallis) confirmed significant differences in fixation rates across semantic categories, with multicollinearity suggesting spatial autocorrelation influenced the distribution of attention (AMATI et al. 2018).

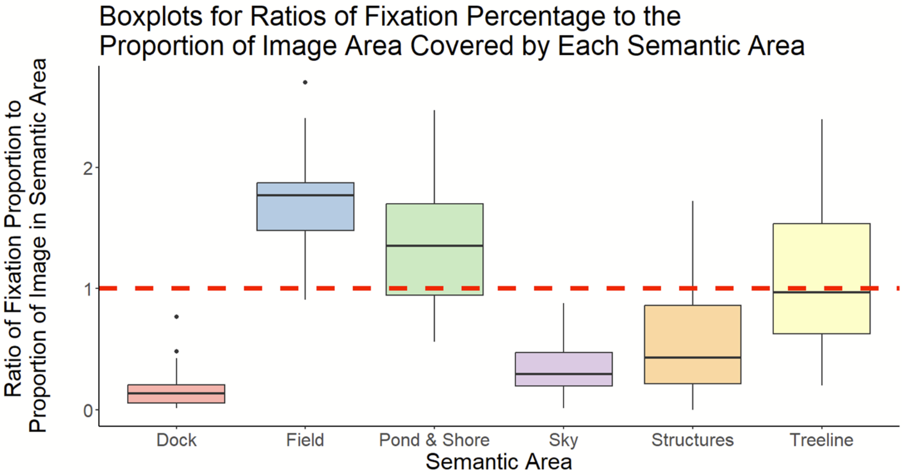

Fig. 4: A boxplot showing the ratio of total fixation points in a semantic area to the percent of the panorama covered by the semantic area

4 Discussion

In this work, we have showcased how eye tracking hardware can be leveraged to assess observations of very large landscapes using AI computer vision approaches. We have provided two representations of the data to explore the data qualitatively (Fig. 3) and quantitatively (Fig. 4). These will be discussed further in this section; first we discuss the broader opportunities and challenges facing eye tracking in large outdoor environments. There are several eye tracking devices available for research, both conventional (fixed head position) and mobile. Our team had access to mobile Tobii Eye Glasses 3. These glasses are limited in depth accuracy for long distances (unreliable beyond a few metres). To overcome this, the gaze points from a single frame can be mapped with high accuracy to a large panorama. In future work, one could consider taking a full panorama and projecting it onto a digital surface model as a texture. This would allow the identification of a fixation in a 3D coordinate space. As in any behavioral experiment involving eye-tracking devices, one potential limitation is that participants may have changed their observational behavior (e. g., RISKO & KINGSTONE 2011), however this may not have been an issue in this study because the observational task did not involve decision making (WORTHY et al. 2024).

The fixation patterns on the transparent heatmap show that participants were focused on the pond's surface and the surrounding reeds, as well as some single trees. These landscape elements are situated in the middle ground of the landscape scene, which can be defined in a visually open scene, characterized by elements such as water and grasslands, as views between 0 and 10 degrees below the horizon (LIU & NIJHUIS 2020). In the scene between 0 - 9 degrees above the horizon, the farm houses caught the fixation of the eye. The importance of the fore- and middle ground for people's visual perception from eye-level perspective has been pointed out by several researchers (APPLETON & LOVETT 2003, LIU & NIJHUIS 2020). Similar to LIU & NIJHUIS (2020), the brightness and distinctiveness of landscape elements seems to attract the viewer's attention. In our study, the pond with the reeds was something participants stated they liked most, while the farm was only mentioned occasionally. Moreover, the *Tilia cordata* tree in the middle ground (and center of the panorama) was mentioned many times as attractive but was not a point of fixation.

This study's free-viewing design leaves participants to visually explore the stimulus according to their interest. The field and pond attracted the largest proportion of attention relative to their size, suggesting a high degree of interest. In contrast, the dock and sky received little attention. The vertical distribution of attention in the scene was clustered towards the middle of the field of view. Bias towards the vertical or horizontal center of a 2D stimulus is commonly noted (MOCHIZUKI et al. 2018) but has rarely been acknowledged for in situ studies. Spatial analysis of gaze data in outdoor environments is limited; methods for addressing center bias and spatial autocorrelation in gaze data require further research to distinguish the effects of an object's significance relative to its size in the field of view.

While we can measure what people observed, future work should explore connections between observational patterns, meaning, and memory. Beyond aesthetic aspects, it is important to consider how scene composition conveys meaning. Future research could examine the relationship between observation and spatial memory using methods like sketch mapping (MANIVANNAN et al. 2024). We must also address movement and noise. Our study controlled for noise with white noise, but this may influence gaze patterns. We included all observations, even those potentially caused by movement (e. g., birds, butterflies, or hikers). Exploring aesthetic value, spatial memory, and movement effects could enhance understanding of design's psychological impacts. Identifying favored landscape elements provides valuable design insights, as COTTET et al. (2018) suggest. They found that natural river sections attract gaze fixations, potentially fostering fascination, well-being, and attention recovery. Stronger links between semantic regions, fixation duration, and cognitive or affective responses are needed to connect these findings to place theory and develop meaningful design indicators (SWITALSKI & GRÊT-REGAMEY 2021).

5 Conclusion and Outlook

Exploring eye-tracking data clustering in outdoor environments reveals insights into visual attention in landscapes. Our study advances the analysis of large outdoor spaces, demonstrating how computer vision enhances eye-tracking data. Our findings support previous research on intriguing phenomena such as the middle ground, brighter objects, water features, and varied unique elements. Future work could integrate these data with psychological factors like spatial memory to better understand how observational patterns shape landscape memory. This approach offers new techniques for landscape architecture to explore how design influences human interaction with landscapes.

Acknowledgements

This research was funded by a Swiss National Science Foundation (SNSF) Research Visit Grant (IZSEZ0_223124), the Swiss State Secretariat for Education, Research and Innovation (SERI), and Eyes4ICU funded by the European Union under the Horizon Europe Marie Skłodowska-Curie Actions, GA No. 101072410. This research was also supported by the Utah Agricultural Experiment Station, Utah State University, and approved as journal paper number #9861. Thank you to Lin Che for discussing the study design and all the participants for their time and dedication.

References

AALTO, P. & STEINERT, M. (2024), Emergence of eye-tracking in architectural research: A review of studies 1976-2021. Architectural Science Review, 1-11.

AMATI, M., PARMEHR, E. G., MCCARTHY, C. & SITA, J. (2018), How eye-catching are natural features when walking through a park? Eye-tracking responses to videos of walks. Urban Forestry & Urban Greening, 31, 67-78. doi:10.1016/j.ufug.2017.12.013.

APPLETON, K. & LOVETT, A. (2003), GIS-based visualization of rural landscapes: defining 'sufficient' realism for environmental decision-making. Landscape and Urban Planning 65 (3), 117-131. doi:10.1016/s0169-2046(02)00245-1.

BRUNS, C. R. & CHAMBERLAIN, B.C. (2019), The influence of landmarks and urban form on cognitive maps using virtual reality. Landscape and Urban Planning, 189, 296-306. doi:10.1016/j.landurbplan.2019.05.006.

CHE, L., RAUBAL, M. & KIEFER, P. (2023), Towards Personalized Pedestrian Route Recommendation Based on Implicit Visual Preference. In Proceedings of the 18th International Conference on Location Based Services, 1-5. doi:10.34726/5639.

COTTET, M., VAUDOR, L., TRONCHÈRE, H., ROUX-MICHOLLET, D., AUGENDRE, M. & BRAULT, V. (2018), Using gaze behavior to gain insights into the impacts of naturalness on city dwellers' perceptions and valuation of a landscape. Journal of Environmental Psychology 60, 9-20. doi:10.1016/j.jenvp.2018.09.001.

DUPONT, L., ANTROP, M. & VAN EETVELDE, V. (2015), Does landscape related expertise influence the visual perception of landscape photographs? Implications for participatory landscape planning and management. Landscape and Urban Planning 141, 68-77. doi.org/10.1016/j.landurbplan.2015.05.003.

HARTRIDGE H. & THOMSON L. (1948), Methods of investigating eye movements. Br. J. Ophthalmol, 32, 581. doi: 10.1136/bjo.32.9.581.

LIU, M. & NIJHUIS, S. (2020), Mapping landscape spaces: Methods for understanding spatial-visual characteristics in landscape design. Environmental Impact Assessment Review, 82, 106376. doi: 10.1016/j.eiar.2020.106376.

LIU, Q., ZHU, Z., ZENG, X., ZHUO, Z., YE, B., FANG, L., ... & LAI, P. (2021). The impact of landscape complexity on preference ratings and eye fixation of various urban green space settings. Urban Forestry & Urban Greening, 66, 127411. https://doi.org/10.1016/j.ufug.2021.127411.

MANIVANNAN C., KRUKAR J. & SCHWERING A. (2024), An algorithmic approach to detect generalization in sketch maps from sketch map alignment. PLoS ONE, 19 (6): e0304696. doi:10.1371/journal.pone.0304696.

MOCHIZUKI, I., TOYOURA, M. & MAO, X. (2018), Visual attention prediction for images with leading line structure. The Visual Computer, 34, 1031-1041. doi.org:10.1007/s00371-018-1518-6.

RISKO, E. F. & KINGSTONE, A. (2011), Eyes wide shut: implied social presence, eye tracking and attention. Attention, Perception, & Psychophysics, 73, 291-296. doi:10.3758/s13414-010-0042-1.

SALVUCCI, D. D. & GOLDBERG, J. H. (2000), Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 symposium on Eye tracking research & applications, 71-78.

SHARMA, S. K., JAIN, K. & SHUKLA, A. K. (2023), A comparative analysis of feature detectors and descriptors for image stitching. Applied Sciences, 13 (10), 6015.

SIEGEL, A.W., WHITE, S.H. (1975), The Development of Spatial Representations of Large-Scale Environments. In: REESE, H. W. (Ed.), Advances in Child Development and Behavior. JAI, 10, 9-55. doi:10.1016/S0065-2407(08)60007-5.

STITES, M., MATZEN, L. & GASTELUM, Z. (2020), Where are we going and where have we been? Examining the effects of maps on spatial learning in an indoor guided navigation task. Cognitive Research: Principles and Implications. 5 (13). doi:10.1186/s41235-020-00213-w.

SWITALSKI, M. & GRÊT-REGAMEY, A. (2021), Operationalising place for land system science. Sustain Sci. 16, 1-11. doi:10.1007/s11625-020-00827-5.

VENABLES, W. N. & RIPLEY, B. D. (2002), Modern Applied Statistics with S. Fourth edition. Springer.

WORTHY, D. A., LAHEY, J. N., PRIESTLEY, S. L. & PALMA, M. A. (2024), An examination of the effects of eye-tracking on behavior in psychology experiments. Behavior Research Methods, 1-14. doi:10.3758/s13428-024-02393-5.