

## MAPPING OF 3D EYE-TRACKING IN URBAN OUTDOOR ENVIRONMENTS

A. Kollert<sup>1</sup>, M. Rutzinger<sup>1</sup>, M. Bremer<sup>1,2</sup>, K. Kaufmann<sup>1,3</sup>, T. Bork-Hüffer<sup>1</sup>

<sup>1</sup> Institute of Geography, University of Innsbruck, Austria – (andreas.kollert, martin.rutzinger, magnus.bremer, katja.kaufmann, tabea.bork-hueffer)@uibk.ac.at

<sup>2</sup> Institute for Interdisciplinary Mountain Research, Austrian Academy of Sciences, Austria – magnus.bremer@oeaw.ac.at

<sup>3</sup> Institute for Comparative Media and Communication Studies, Austrian Academy of Sciences and University of Klagenfurt, Austria – katja.kaufmann@oeaw.ac.at

### Commission II, WG II/10

**KEY WORDS:** Eye-tracking, Heatmap, 3D Model, Outdoor Environment, Photogrammetry, Geospatial Digital Humanities

### ABSTRACT:

New geospatial technologies and ubiquitous sensing allow new insights into people's spatial practices and experiences of public spaces. These tools offer new data streams for analysis and interpretation of social phenomena. Mobile augmented reality tools such as smartphones and wearables merge the experience of entangled online and offline spaces in citizen's daily life. This paper demonstrates a concept that combines eye-tracking tools with innovative mapping in order to enhance the interpretability of real outdoor environmental experiences. Through videogrammetry, a participants' head posture can be reconstructed. Subsequently the fixations measured through eye-tracking are projected onto a 3D point cloud of the surrounding environment. The presented methodological approach is implemented in the interdisciplinary project DigitAS – The Digital, Affects and Space – which investigates the perception of public places as spaces of recreation, security or fear. The project's Mixed Methods approach combined qualitative, mobile, in-situ and reconstructive methods with eye-tracking in an outdoor setting. Potentials of the geospatial mapping concept for social science research is discussed.

## 1. INTRODUCTION

### 1.1 The Digital, Affects and Space

People's involvement with mobile social media and emerging technologies such as virtual and augmented reality is manifold and increases the complexity of their perception of the world (Dey et al., 2018; Felgenhauer and Gäbler 2018; Lemos 2008; Malpas 2008). Indeed, these digital technologies are subtly co-producing social life and entangled socio-material-technological spaces (Kitchin and Dodge, 2011; Sumartojo et al., 2016; Bork-Hüffer et al., 2020). In face of accompanying digital phenomena such as hate speech, trolling or so-called "fake news", researching the implications of using digital technologies seems more urgent than ever. Yet, due to methodological limitations, attempts to analyse the effects of the in-situ use of digital media, especially augmented reality, on people's affective-emotional experiences of space are still limited. Affective-emotional perception is explicitly defined as including the continuum from pre-cognitive, subconscious to conscious and communicated bodily and sensuous experiences (Schurr and Strüver 2016). Augmented Reality (AR) is a field of technologies which imports digital images in real world settings "in such a way that the virtual content is aligned with real world objects, and can be viewed and interacted with in real time" (Dey et al., 2018, 1). Against this background, the interdisciplinary mixed methods development project "The Digital, Affects and Space" (DigitAS) was set up. Core of the project was the implementation of a quasi-experimental field study, which combined mobile eye-tracking with subsequent retrospective think-alouds (Konrad, 2010) to research people's in-situ use of digital augmented media in public parks and its influence on their perception of such places.

### 1.2 Mapping Concept

The feasibility of 3D reconstruction of the head pose based on eye-tracking data in urban outdoor environments is investigated using Structure-from-Motion (SfM). While several studies have targeted 3D-gaze mapping based on eye-tracking systems in indoor environments (Hagihara et al., 2018; Matsumoto and Takemura, 2019; Maurus et al., 2014; Wang et al., 2018), studies that apply similar techniques in urban outdoor environments are yet rare. Successful reconstruction of the 3D head position however is crucial for mapping 3D gaze positions. Therefore, our results will be valuable for eye-tracking practitioners that pursue 3D gaze mapping. Currently no end-to-end solution for such a task exists (Kiefer et al., 2017). Even if not pursuing 3D gaze mapping, a reconstruction of the eye-tracking frames helps, if successful, to achieve a higher positional accuracy than possible with consumer grade Global Navigation Satellite System (GNSS) solutions.

### 1.3 Mobile eye-tracking

In Mixed Methods Research (MMR), combinations including mobile eye-tracking have increasingly been applied for a wide variety of different research objects (e.g., Wang and Sparks, 2016; Hoy and Levenhus, 2018; Jankowska et al., 2018; Stark et al., 2018). Yet, so far, MMR designs using eye-tracking are usually applied in lab conditions, while field experiments (e.g., Vater et al., 2019) are comparatively rare. But the technological improvement of mobile eye-trackers increasingly allows the use of this instrument in diverse real-world settings and thus also for geo-applications: Mobile eye-tracking is used for capturing the perception of street edges (Simpson et al., 2019), to

investigate human navigation (Liao et al., 2019), self-localization (Kiefer et al., 2014) and more generally, the perception of urban environments (Hollander et al., 2019; Sussman and Ward, 2019). Since all of these studies employ mobile eye-tracking devices, they all have a spatial component. Kiefer et al. (2012) introduced the concept of *location aware mobile eye-tracking* in order to explicitly emphasize the use of 2D and 3D positional data, i.e. spatial data, in eye-tracking experiments. As Kiefer et al. (2012) note, many studies have evaluated how people perceive and interact with spatial data, but rarely eye-tracking data and positional data have been combined. With the exception of Kiefer et al. (2014), none of the examples given here employ positional data in the way Kiefer et al. (2012) have described. More recently, Tomasi et al. (2016) have introduced a method to capture head orientation with respect to body position. Their case study underlines the relevance of head movements in outdoor eye-tracking, which is also in line with Singh et al. (2018), who found that head centric sampling might be suitable to approximate gaze. Still, studies that integrate eye-tracking and positional data, in order to allow for further insights, remain rare. 3D reconstruction (see Section 1.4), however, could stimulate the use of positional data in mobile eye-tracking, since reconstruction of position and orientation is inherent to these methods.

#### 1.4 3D Gaze mapping

Matsumoto and Takemura (2019) have presented a framework for 3D gaze mapping using SfM without the need to obtain a 3D model prior to eye-tracking. They prove that this works well for a spatially confined indoor environment using only imagery recorded by the eye-tracker. Singh et al. (2018) build a solution for 3D gaze mapping in outdoor environments by employing similar methods as Matsumoto and Takemura (2019). Their method is demonstrated on a textured statue model and fixations are visualized on this model. Singh et al. (2018) point out that eye-tracking data can be supplemented by high quality imagery due to the low image quality of most eye-tracking devices. A similar workflow is used by Li et al. (2020) who also employ SfM on video footage acquired by a mobile eye-tracking device. Hagihara et al. (2018) introduce a hybrid approach, where 3D models are computed based on a depth camera, but the exact trajectory of a person is computed based on SfM. Subsequently both data sets are co-registered.

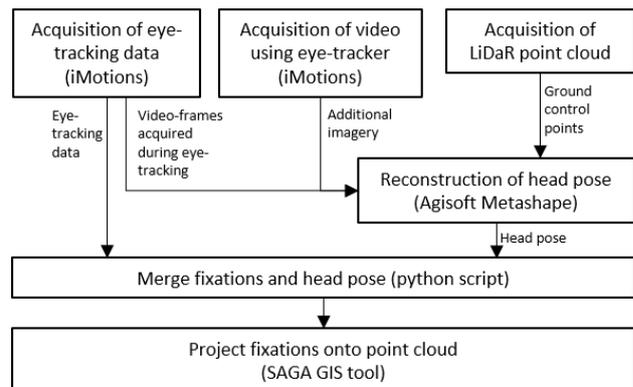
Although not specifically targeted at such, the approaches by Hagihara et al. (2018), Matsumoto and Takemura (2019) and Li et al. (2020) are presented with indoor applications. Only Singh et al. (2018) explicitly show an outdoor scene. This of course raises the question of how well such concepts can be transferred into other environments where environmental conditions (e.g. illumination) are subject to change and cannot easily be controlled or cannot be controlled at all. This point was also made by Singh et al. (2018) who stated that such an effort could be challenging. Reconstructing the external orientation of every video frame however is crucial for subsequent gaze mapping because the recording rate of the eye-tracking system is usually higher than the recording rate of the scene camera.

#### 1.5 Aim of the paper

The aim of this paper is to investigate the 3D reconstruction of eye-tracking data based on a case study. In line with previous approaches, we employ SfM for reconstructing the participants' paths through a park and project their recorded fixations back onto a 3D model using the image geometry derived by SfM. We seek to investigate the applicability, ease of implementation and

necessary ancillary data sources in this process. In order to be able to georeference a photogrammetric model, we rely on a 3D model based on a terrestrial laser scan. The eye-tracking videos are not optimized for being integrated into Dense Matching and 3D model reconstruction, i.e. imagery is acquired using the built-in sensor of the eye-tracking system.

## 2. METHODS



**Figure 1.** Workflow for 3D reconstruction and projection of fixations.

### 2.1 Study site

The study was conducted in the Rapoldi Park in Innsbruck, Austria (47° 15' 57.9" N, 11° 24' 22.5" E). Figure 2 shows the area of interest. While participants of the study had to follow a specific path through the park (Figure 2, red line), only a subset of this route was selected for the 3D reconstruction of fixations as illustrated by the blue polygon. This subset of the route amounts to about 50 m in length at most.

In the framework of the DigitAS-project, a larger eye-tracking study was conducted within this specific location. In order to explore the potential of these recordings for 3D reconstruction and mapping, a subset of this site was chosen for implementing the workflow (Figure. 1).



**Figure 2.** Study site with entire route for participants of the study and sub-area for 3D reconstruction of fixations.

### 2.2 Eye-tracker & Camera specifications

Tobii Pro Glasses 2 were used for capturing gaze data at 50 Hz. A wide-angle scene camera captures the environment at 25 frames per second (fps), i.e. half the rate of the gaze data. The resolution of the scene camera is 1920x1080 pixels (approx. 2 MP) and its field-of-view amounts to 82° (horizontal) and 52° (vertical). The scene camera has a focal length of 3.36 mm and a sensor size of 2.7328 mm x 1.5344 mm (Tobii Pro AB, 2021).

The recordings of the scene camera are used to visualize the participant's gaze within iMotions software (<https://imotions.com/>, iMotions, 2018) and represent the basis

for the production of 2D and 3D heatmaps. During recording, all data is stored on a local Secure Digital (SD) card within the recording unit carried by the participant. At the end of each experiment the data is transferred to a workstation for further analysis of the gaze data. Fixations were calculated within iMotions using the I-VT Fixation Filter with default parameters. For details on these parameters, the reader is referred to Olsen and Matos (2012). 2D heatmaps were produced within iMotions. For external processing, eye-tracking data was exported from iMotions.

### 2.3 Eye-tracking case study

During the DigitAS study, which took place in September and October 2020, 10 participants were given the task to walk a pre-defined route through the park twice while wearing the eye-tracking device (Figure 2). Participants did not have prior knowledge about the route and environment beforehand. For the first walk, no specific task was given to the participants other than walking the specified route. For the second walk, participants were given a smartphone. At certain locations during this second walk, participants received multimedia content with information about this location using an instant messenger. Such a message could, for example, contain a media report on the newly built playground or a forum post on criminal activities that had happened just at this location of the park. Participants were asked to pause their walk when receiving such a message, read the content and continue the walk afterwards. The aim of this procedure is to investigate the impact of in situ media consumption on the perception of public places via eye-tracking (c.f. Kaufmann et al. 2021). For each participant, two eye-tracking data sets are available: One data set collected during the first (“regular”) walk and one data set that relates to the second walk were additional location-specific information was given to the participant.

Out of 10 participants, one participant was selected to be included in the 3D processing workflow presented in this paper. Since the aim is to investigate the feasibility of our workflow design, this selection was mainly guided by data quality, i.e. good coverage of the eye-tracker. In some cases and areas, eye-tracking coverage was very low, presumably due to strong direct sunlight. Restricting our 3D-processing to one participant only was necessary due to the relatively high manual workload.

Walking the entire route took on average 8 minutes for each of the participants. For the selected participant and each of their two walks, we selected a subset that covered our area of interest (Figure 2). This resulted in snippets of 20 to 40 seconds of video and eye-tracking data, depending on how much time the participant would spend in that area. 2D and 3D heatmaps were computed for each of these two subsets.

### 2.4 Acquisition of additional data for 3D model

A terrestrial laser scanning (TLS) campaign was conducted in order to obtain a sound 3D model of our area-of-interest (AOI, see Figure 4). Later in the process, the fixations are mapped onto this 3D point cloud by ray-tracing. Furthermore, Ground-Control-Points (GCPs) for georeferencing our photogrammetric model are derived from this point cloud. The point cloud was acquired using a Riegl VZ-2000i TLS. For 3D mapping of the fixations (see Section 2.7), a thinned-out version of this point cloud was used.

During recording of the eye-tracking data, no considerations were given to the derivation of image geometry later by SfM. Initial attempts to reconstruct the trajectory of participants (see Section 2.6) using only this set of imagery failed. This can likely be attributed to the relatively low image quality, e.g. no possibility to manually adjust or fix any camera parameters of the eye-tracker and weak geometry of the scene (Singh et al., 2018). Therefore, an auxiliary acquisition was conducted. During this auxiliary acquisition additional imagery for the 3D trajectory reconstruction was collected along and around the track of the eye-tracking acquisitions using the same eye-tracking glasses, in order to create a stable photogrammetric block. Since the amount of video frames that need to be processed quickly goes into the thousands, the area for which additional imagery was collected needed to be restricted (see Figure 2). This additional imagery was collected by mounting the eye-tracker onto a longer pole which enabled coverage the participant’s track and the surroundings and provided oblique views of the scene. This smaller sub-area for which we collected auxiliary imagery then became the AOI for 3D reconstruction of the fixations.

### 2.5 Postprocessing of the eye-tracking and video data

2D heatmaps (Figure 5a and Figure 5b) were calculated within iMotions (iMotions, 2018). The 2D heatmaps from iMotions as displayed in this paper offer no absolute or quantitative information, the color scale illustrates the relative attention within a scene. Red implies that a certain region within an image has been visually fixated for a longer time, yellow and green areas imply medium to little time spent in an area. Areas without any color have not been fixated at all (iMotions, 2021). The image or video frame that is used for a 2D heatmap has to be specified by the user. The same applies to the time span that is considered for mapping fixations onto that specified image.

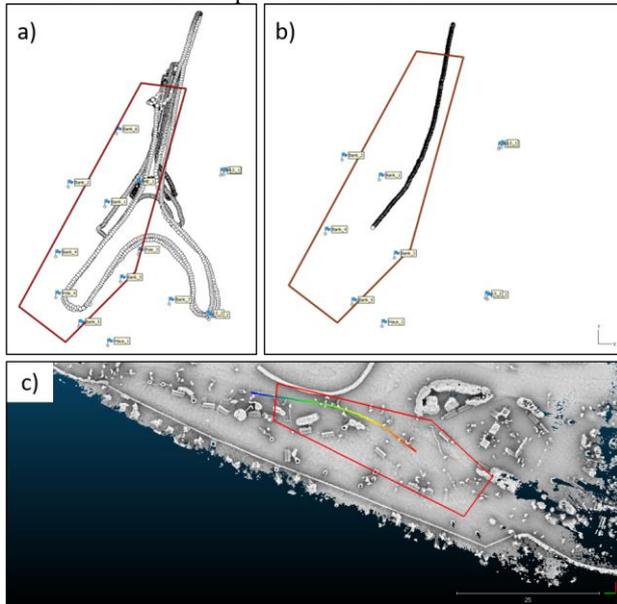
For 3D processing, eye-tracking data was exported from iMotions. Fixations were stored as image coordinates for each video frame. Single frames were extracted from the video of the scene camera that is simultaneously captured during operation of the eye-tracking glasses. In order to be able to temporally align the images to the eye-tracking data later on, the frame-number was written to individual video frames. Thereby the timestamp of each frame can later be reconstructed using the frame rate. The frame rate nominally amounts to 25 fps, the actual value to be found in the metadata of the video however is 25.01 fps.

### 2.6 3D head pose estimation

In order to establish the exact position and orientation of the images captured by the scene camera of the eye-tracker, a photogrammetry-based workflow was employed. The idea is to embed the imagery collected during eye-tracking of individuals into a greater photogrammetric block of images, which has in parts been presented by Singh et al. (2018) and others before (see Section 1.4). Since a point cloud of the scene has been acquired before, there is no need to perform a dense reconstruction of the scene, but only to perform SfM to obtain the image geometry. This is sufficient to allow for the 3D mapping of the participant’s gaze or fixation onto the 3D point cloud. In essence, our workflow includes the regular SfM pipeline.

For imagery from the first and second walk of the selected participant the image geometry was reconstructed. Individual video frames were extracted from the recordings of the scene

camera at full rate (25.01 fps). Since the eye-tracking snippets of the two walks amount to 25 s and 45 s, this resulted in approx. 500 and 1000 video frames, respectively, for the eye-tracking study. Auxiliary imagery collected on a second date with the same sensor provided another 2500 images. All of the images were imported to Agisoft Metashape v1.6.6 (Agisoft, 2020) for processing. In a first step, image alignment was carried out using the accuracy setting “high” with a key point limit of 40000 and a tie point limit of 4000.



**Figure 3.** a) White dots show the camera positions of additional imagery as reconstructed within Agisoft Metashape, b) dots show the camera positions of the eye-tracking imagery, i.e. the head pose of the participant, during the first walk and c) illustrates the point cloud acquired with a terrestrial laser scanner. The red polygon in all of the subfigures here corresponds to the AOI for 3D reconstruction in Figure 2.

In order to obtain a georeferenced model, synthetic GCPs were extracted from the TLS point cloud. Within the AOI, artificial objects such as corners of benches, poles and objects within the playground were used as GCPs. The GCPs are relatively well distributed horizontally over the small AOI. Figure 4 shows the GCPs that were used for georeferencing the model (red points and labels).



**Figure 4.** GCPs used to geo-reference the photogrammetric model based on the eye-tracking recordings.

After reconstruction of the image geometry without GCPs, the automatic marker detection in Metashape was run and the GCP positions were manually refined if necessary. In an earlier stage, the camera was calibrated and this calibration was used throughout the entire processing in Agisoft Metashape. After bundle block adjustment, the root mean square error for all

GCPs amounts to 0.305 m. Although video frames from eye-tracking and auxiliary imagery were processed together, the reconstructed image geometry for both is illustrated in separate figures (Figure 3a and Figure 3b).

## 2.7 Mapping of fixations onto point cloud

After successful computation of the exterior orientation of the eye-tracking frames, the frames were merged with the eye-tracking data based on the timestamp. Considering exterior and interior camera orientations, fixations were transferred from image coordinates into 3D direction vectors and then mapped onto the point cloud using a ray-tracing scheme. This scheme was implemented as a SAGA GIS tool (Conrad et al., 2015). Within this scheme, the point cloud is partitioned into voxels of 10 cm size. The ray-tracing was not only performed for the beam of a single fixation, but for multiple rays within a specified opening angle around the direction vector. For our purpose, this angle was set to 5°. If a ray hits one of the voxels, a counter is incremented. After iterating over all fixations, each voxel cell has a certain count, which can then be displayed as 3D heatmap.

## 3. RESULTS

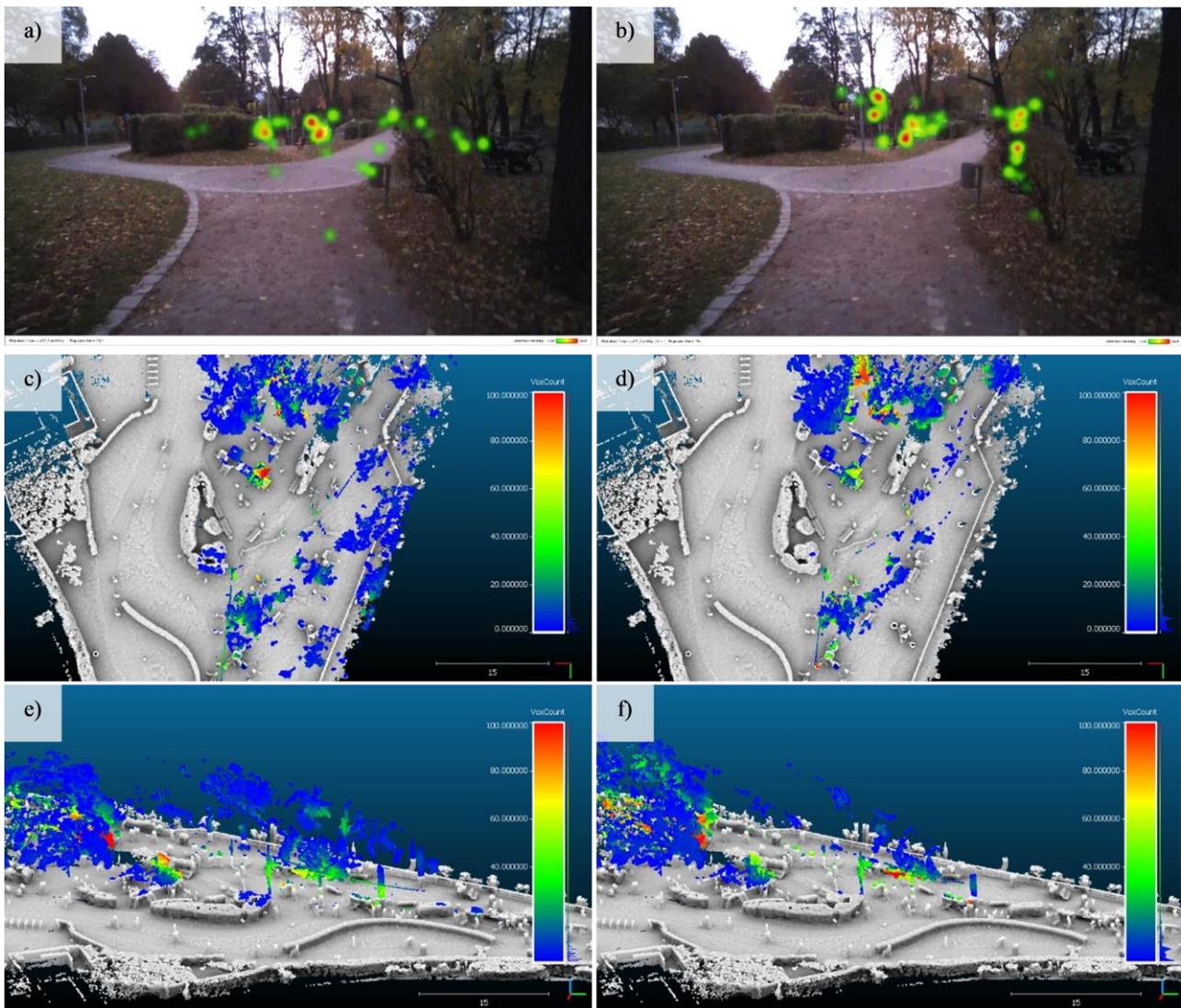
### 3.1 2D heatmaps of the scene

Regular two-dimensional heatmaps were exported directly from iMotions for comparison with the three-dimensional reconstruction. Figure 5a shows the heatmap for the first walk of the selected participant based on the fixations. The heatmap was produced using iMotions standard settings, i.e. no adjustments were made. This 2D heatmap shows that the participant devotes some attention to the bushes on the right side of the walk, but most of their attention is devoted to the central part in the image where a small playground is located.

Figure 5b shows the heatmap for the second walk of the participant mapped onto the same video frame. The overall pattern is similar to the first walk (Figure 5a), however the attention in the central part of the image is located more towards the top of the image. Generally, one has to keep in mind, that for computing the heatmaps, the time span was shorter in Figure 5a (25 s) compared to Figure 5b (45 s). The section and therefore also length of the walk, however, was chosen to roughly be the same in both cases.

### 3.2 3D reconstruction of participants fixations

Figure 5 (c - f) show fixations that were mapped onto the point cloud for the first walk. Figure 5 (c, d) and Figure 5 (e, f) illustrate the view from the top and left side of the walk, respectively. Red color indicates that many fixations hit a specific point, whereas dark blue indicates only few fixations. When compared to Figure 5a and Figure 5b (2D heatmaps), the 3D representation of the same temporal subset of eye-tracking data gives quite a different representation of the entire scene. Of course, this can also be traced back to differences in the methodology of producing these two representations, but the 3D model offers the possibility to be inspected from every angle, for every point or facet, whereas the 2D heatmap is quite limited in the way it can be explored.



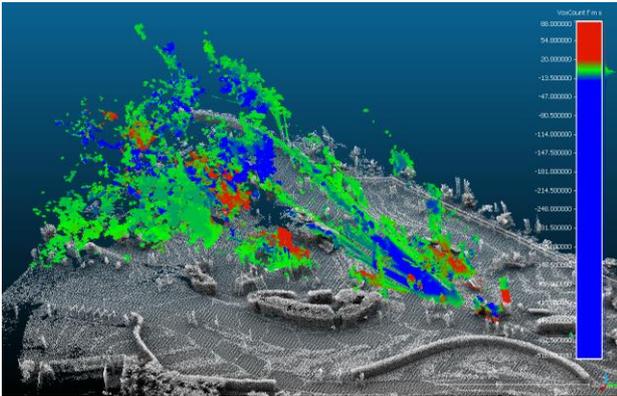
**Figure 5.** 2D and 3D heatmaps for the selected participant. Subfigures in left column (a, c, e) and right column (b, d, f) are from the first and second walk, respectively. a) and b) are 2D heatmaps of the scene produced with iMotions. Red color indicates areas that have been fixated for a longer time, while yellow to green indicate medium to little attention. c), d), e) and f) are representations of our 3D heatmaps of the very same scene and eye-tracking recordings: The grey point cloud is shown for reference only; elements with a height above ground greater than 1.5 m were filtered out in order to exclude vegetation that might obscure the scene. Points that have been hit by fixations (see Section 2.7 for explanation) are shown red to blue. Again, red indicates a high number of fixations, yellow to green to blue indicate medium to few fixations. Points that have not been hit by a fixation are not shown, except for the grey background model.

By comparing Figure 5c and Figure 5d, it is possible to analyse how the spatial distribution of the participant's attention changes in different scenarios. Figure 5c from the first walk for example shows that the participant allocated more attention towards the right side of the picture (dark blue spots). This is also obvious when comparing Figure 5e and Figure 5f: Figure 5e shows many points in the upper part of the figure, that are not visible at all in Figure 5f. These points are not visible in Figure 5f because they have not been reached by any fixations of the participant. One can identify a red spot in the center of Figure 5f which is not that pronounced in Figure 5e. At this spot a bench as well as some bushes are located. While a similar pattern can be identified in Figure 5b, which is the 2D heatmap complementary to Figure 5d and Figure 5f, the 3D model offers more detail and insight, since it can be inspected in more dimensions and detail and spatially relating the topographic situation, objects and the participant's location to each other.

Using the 3D heatmaps, which are actually stored as an attribute of the underlying point cloud, it is also possible to compute the difference between the two walks of the participant. Figure 6 illustrates this difference. Here, the roof of the playground in the central part of the image attracted more fixations during the first walk. In the second walk, the participant devoted more attention towards the path itself, but also the surrounding upper part of the playground, as indicated by the dark blue color.

#### 4. CONCLUSION & OUTLOOK

We have shown that the position of participants during eye-tracking can be successfully reconstructed in outdoor environments. Despite the availability of photogrammetric software solutions, results indicate that 3D reconstruction based on the scene recordings from eye-tracking glasses can fail if only relying on the video frames acquired during eye-tracking.



**Figure 6.** Difference of fixation count for each point, computed as number of fixations of the first walk minus number of fixations of the second walk. Red color indicates that more fixations have occurred during the first walk, while blue color indicates that more fixations occurred during the second walk. Green color illustrates the same level of fixations in both walks

Thus, additional image collection and manual refinements during the 3D reconstruction were still necessary. In our case, an accurate 3D model acquired with a terrestrial laser scanner was necessary in order to allow for georeferencing of the model obtained within Agisoft Metashape. So far, recent studies that employ similar methods have specifically targeted indoor scenes. We were able to show the successful combination and application of photogrammetric mapping methods with an eye-tracker in urban outdoor environments.

Now that we have explored the feasibility of 3D eye-tracking reconstruction in outdoor environments, the proposed methods are applied within the DigitAS project for the analysis of eye-tracking data collected in public parks. Thereby the methods will help to understand how media consumption affects the perception of public spaces by individuals. Based on our initial results, tools will be developed that allow mapping of gaze data onto a 3D representation of the environment, e.g. TLS point clouds or mesh data. Future work should focus on the real-time processing of video stream data, eye-tracking, and object detection integrating positioning data streams. An accuracy assessment of the 3D eye-tracking reconstruction is still outstanding as well and it remains unclear if such a step can be integrated in a case study such as ours without impairing the study design itself (e.g. if additional sensors are required). Within social sciences, endeavours to analyse social reality while it unfolds in-situ have increased throughout the last 15 years by means of so-called mobile methods (cf., e.g., Boase and Humphreys 2018, Moles 2019). Eye-tracking represents a fruitful approach to investigate subtle experiences of space that humans are less aware of and less able to verbally express as part of traditional research methods such as interviews. Yet, visualisations of eye-tracking data remain restricted so far as 2D visualisations merely reflect subjects' position in space but not their actual engagement with their material surroundings. The proposed 3D mapping concept bridges this gap. Mapping fixations in 3D space also shows potential for situations where participants move a lot and fixations are tracked over a longer time span. In such a case, it would be necessary to compute different 2D heatmaps for different views of scene, whereas a 3D representation offers all the information within a single model. Still, currently, work involved in producing the 3D eye-tracking maps is quite time-intensive and automatization of data integration is needed to make the maps of interest to the interpretation of social phenomena beyond a singular case.

Nonetheless, 3D visualizations can be used to communicate the project's empirical findings to stakeholders that are familiar with geospatial data, such as landscapers and city planners to assess the perception of the public park. Additional research is needed to investigate to what an extent these 3D heatmaps can enable participants to verbalize their subjective experiences while wearing the mobile eye-tracker more adequately or fully compared to 2D heatmaps as well as recorded videos in retrospection elicitation.

## ACKNOWLEDGEMENTS

This work has been undertaken in the project DigitAS - The Digital, Affects and Space (<http://digitas.mountainresearch.at/>) funded by the program GO!DIGITAL of the Austrian Academy of Sciences promoting research in digital humanities. Furthermore, we thank the participants for taking part in our study.

## REFERENCES

- Agisoft Metashape Professional (Version 1.6.6) (Software). (2020). Retrieved from <http://www.agisoft.com/downloads/installer/> (Accessed 22 January, 2021)
- Boase, J., Humphreys, L. (2018). Mobile methods: Explorations, innovations, and reflections. *Mobile Media & Communication*, 6, 153-162. doi.org/0.1177/2050157918764215
- Bork-Hüffer, T., Mahlknecht, B., Markl, A. (in print). 8 Thesen zu Verschränkungen, multiplen Historizitäten und Intra-Aktionen in sozio-materiell-technologischen (Alltags-)Räumen“ [8 propositions on entanglements, multiple historicities, and intra-actions in socio-material-technological (everyday) spaces]. *Zeitschrift für Kultur- und Kollektivwissenschaften*.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific Model Development*, 8(7), 1991-2007. doi.org/10.5194/gmd-8-1991-2015
- Dey, A., Billingham, M., Lindeman, R. W., Swan, J. (2018). “A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014”, *Frontiers in Robotics and AI*, Vol. 5 No. 37. doi.org/10.3389/frobt.2018.00037
- Felgenhauer, T., Gäbler, K. (2018). *Geographies of digital culture: an introduction*. Abingdon, New York, Routledge.
- Hagihara, K., Taniguchi, K., Abibouraguimane, I., Itoh, Y., Higuchi, K., Otsuka, J., Sugimoto, M., Sato, Y. (2018, February). Object-wise 3D Gaze Mapping in Physical Workspace. In *Proceedings of the 9th Augmented Human International Conference* (pp. 1-5). doi.org/10.1145/3174910.3174921
- Hollander, J. B., Purdy, A., Wiley, A., Foster, V., Jacob, R. J., Taylor, H. A., Brunyé, T. T. (2019). Seeing the city: using eye-tracking technology to explore cognitive responses to the built environment. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 12(2), 156-171. doi.org/10.1080/17549175.2018.1531908

- Hoy, M. G., Levenhus, A. B. (2018). A mixed-methods approach to assessing actual risk readership on branded drug websites. *Journal of risk research*, Vol. 21 No. 5, pp. 521-538. doi.org/10.1080/13669877.2016.1223160
- iMotions Biometric Research Platform (v8.2), iMotions A/S, Copenhagen, Denmark, 2018.
- iMotions (2021). Basic Heatmap Explanation and Generation. <https://help.imotions.com/hc/en-us/articles/360011496079-Basic-Heatmap-Explanation-and-Generation> (Accessed 22 January, 2021)
- Jankowska, D. M., Czerwonka, M., Lebuda, I., Karwowski, M. (2018). Exploring the Creative Process: Integrating Psychometric and Eye-Tracking Approaches. *Frontiers in psychology*, Vol. 9, pp. 1931-1931. doi.org/10.3389/fpsyg.2018.01931
- Kaufmann, K., Bork-Hüffer, T., Gudowsky-Blatakes, N., Rauhala, M., Rutzinger, M. (2021). Ethical Challenges of Researching Emergent Socio-Material-Technological Phenomena: Insights from an Interdisciplinary Mixed Methods Project Using Mobile Eye-Tracking. *Journal of Information, Communication and Ethics in Society*, in press.
- Kiefer, P., Giannopoulos, I., Raubal, M. (2014). Where am I? Investigating map matching during self-localization with mobile eye tracking in an urban environment. *Transactions in GIS*, 18(5), 660-686. doi.org/10.1111/tgis.12067
- Kiefer, P., Giannopoulos, I., Raubal, M., Duchowski, A. (2017). Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation*, 17(1-2), 1-19. doi.org/10.1080/13875868.2016.1254634
- Kiefer, P., Straub, F., Raubal, M. (2012, March). Towards location-aware mobile eye tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 313-316). doi.org/10.1145/2168556.2168624
- Kitchin, R., Dodge, M. (2011). *Code/space: software and everyday life*. London, Cambridge: MIT Press.
- Konrad, K. (2010). "Lautes Denken" [Thinking aloud]. In: Mey, G., Mruck, K. (Ed.s), *Handbuch Qualitative Forschung in der Psychologie*, Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 476-490.
- Lemos, A. (2008). Mobile communication and new sense of places: a critique of spatialization in cyberculture. *Galaxia*, Vol. 16, pp. 91-109.
- Liao, H., Dong, W., Huang, H., Gartner, G., Liu, H. (2019). Inferring user tasks in pedestrian navigation from eye movement data in real-world environments. *International Journal of Geographical Information Science*, 33(4), 739-763. doi.org/10.1080/13658816.2018.1482554
- Li, T. H., Suzuki, H., Ohtake, Y. (2020). Visualization of user's attention on objects in 3D environment using only eye tracking glasses. *Journal of Computational Design and Engineering*, 7(2), 228-237. doi.org/10.1093/jcde/qwaa019
- Malpas, J. (2008). New media, cultural heritage and the sense of place: Mapping the conceptual ground. *International Journal of Heritage Studies*, Vol. 14, pp. 197-209. doi.org/10.1080/13527250801953652
- Matsumoto, R., Takemura, K. (2019, October). Semantic 3D gaze mapping for estimating focused objects. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-6). doi.org/10.1145/3338286.3344396
- Maurus, M., Hammer, J. H., Beyerer, J. (2014, March). Realistic heatmap visualization for interactive analysis of 3D gaze data. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 295-298). doi.org/10.1145/2578153.2578204
- Moles, K. (2019). Mobile Methods. In: Atkinson, P. A., Delamont, S., Williams, R. A., Cernat, A., Sakshaug, J. (eds.) *SAGE Research Methods Foundations*. doi.org/10.4135/9781526421036847435
- Olsen, A., Matos, R. (2012, March). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In *proceedings of the symposium on Eye tracking research and applications* (pp. 317-320). doi.org/10.1145/2168556.2168625
- Schurr, C., Strüver, A. (2016). "The Rest": Geographien des Alltäglichen zwischen Affekt, Emotion und Repräsentation. *Geographica Helvetica*, 71, 87-97. doi.org/10.5194/gh-71-87-2016
- Simpson, J., Freeth, M., Simpson, K. J., Thwaites, K. (2019). Visual engagement with urban street edges: insights using mobile eye-tracking. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 12(3), 259-278. doi.org/10.1080/17549175.2018.1552884
- Singh, K., Kalash, M., Bruce, N. (2018, June). Capturing real-world gaze behaviour: live and unplugged. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (pp. 1-9). doi.org/10.1145/3204493.3204528
- Stark, L., Brünken, R., Park, B. (2018). Emotional text design in multimedia learning: A mixed-methods study using eye tracking. *Computers and education*, Vol. 120, pp. 185-196. doi.org/10.1016/j.compedu.2018.02.003
- Sumartojo, S., Pink, S., Lupton, D., LaBond, C. H. (2016). The affective intensities of datafied space. *Emotion, Space and Society*, Vol. 21, pp. 33-40. doi:10.1016/j.emospa.2016.10.004
- Sussman, A., Ward, J.C. (2019). Eye-tracking Boston City Hall to better understand human perception and the architectural experience. *New design ideas*, 3(1), 53-59.
- Takemura, K., Takahashi, K., Takamatsu, J., Ogasawara, T. (2014). Estimating 3-D point-of-regard in a real environment using a head-mounted eye-tracking system. *IEEE Transactions on Human-Machine Systems*, 44(4), 531-536. doi.org/10.1109/THMS.2014.2318324
- Tobii Pro AB (2021). Tobii Pro Glasses 2. Tobii Pro, Danderyd, Sweden. <https://www.tobiiipro.com/de/produkte/tobii-pro-glasses-2/>. (Accessed 21 January 2021)

Tomasi, M., Pundlik, S., Bowers, A. R., Peli, E., Luo, G. (2016). Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of vision*, 16(3), 27-27. [doi.org/10.1167/16.3.27](https://doi.org/10.1167/16.3.27)

Vater, C., Luginbühl, S., Magnaguagno, L. (2019). Testing the functionality of peripheral vision in a mixed-methods football field study. *Journal of sports sciences*, Vol. 37, No. 24, pp. 2789-2797. [doi.org/10.1080/02640414.2019.1664100](https://doi.org/10.1080/02640414.2019.1664100)

Wang, H., Pi, J., Qin, T., Shen, S., Shi, B. E. (2018, June). SLAM-based localization of 3D gaze using a mobile eye tracker. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (pp. 1-5). [doi.org/10.1145/3204493.3204584](https://doi.org/10.1145/3204493.3204584)

Wang, Y., Sparks, B. A. (2016). An Eye-Tracking Study of Tourism Photo Stimuli: Image Characteristics and Ethnicity. *Journal of Travel Research*, Vol. 55 No. 5, pp. 588-602. [doi.org/10.1177/0047287514564598](https://doi.org/10.1177/0047287514564598)