Using Eye Tracking Data to Understand Visitors' Behaviour

Marco Grazioso
Department of Electrical Engineering and Information Technology, University of Naples 'Federico II'
Naples, Italy
marco.grazioso@unina.it

Roberto Esposito
Department of Electrical Engineering and Information Technology, University of Naples 'Federico II'
Naples, Italy
zanroby@live.it

Emma Maayan-Fanar
Art History Department, University of Haifa
Haifa, Israel
efanar@univ.haifa.ac.il

Tsvi Kuflik
Information Systems Department, University of Haifa
Haifa, Israel
tsvikak@is.haifa.ac.il

Francesco Cutugno
Department of Electrical Engineering and Information Technology, University of Naples 'Federico II'
Naples, Italy
cutugno@unina.it

ABSTRACT

In the last decade, mobile eye trackers have become a commodity. With the decrease in their cost and their increased availability, a growing number of studies are exploring their potential in cultural heritage. In works on art fruition, the idea of using mobile eye trackers to understand how we observe artistic exhibits is becoming commonplace. Following this trend, the goal of this study was to explore the potential and to propose an eye-tracking method to analyse the behaviour of museum visitors while they observe artworks, using the Hecht Museum in Haifa as a case study. Recorded data from two different observation sessions, before and after a course in art history, were used to produce heat maps, which serve as a simple and effective tool for studying changes in visitors' gaze patterns when observing artworks.

CCS CONCEPTS

• **Human-centered computing → Heat maps**; • **Applied computing → Arts and humanities**.

KEYWORDS

Mobile Eye-Tracker, Art Appreciation, Data Analysis, User Behaviour

ACM Reference Format:

Marco Grazioso, Roberto Esposito, Emma Maayan-Fanar, Tsvi Kuflik, and Francesco Cutugno. 2020. Using Eye Tracking Data to Understand Visitors' Behaviour. In *Proceedings of AVI²CH 2020: Workshop on Advanced Visual Interfaces and Interactions in Cultural Heritage (AVI²CH 2020)*. ACM, New York, NY, USA, 7 pages.

1 INTRODUCTION

How people look at works of art involves a mixture of individual factors ranging from cultural background to personal sensibility. Thus there is no "correct" way to look at art. Works of art might convey different meanings and transmit different emotions and feelings to different observers. Nevertheless, we also know that artists try to instill their ideas in their work, as is also the case, for example, in architecture [16], where the architect's ideas are expressed, implicitly or explicitly, in the building design. Artists, thus, have a message they wish to convey or an idea they are trying to express through their work. If we fail to pay attention to the details the artist wants us to focus on, or fail to discern how the different parts of the image cohere, we will probably miss the message or the idea the artist is trying to convey [3].

According to Rorty [15], works of art typically express the prevailing theological, political, social, and economic beliefs of their time but we need not assume that these are what the painting is *about*. Whether direct or indirect, conscious or unconscious, these beliefs serve merely as pointers to or questions about the visual experience of the painting, which cannot be conveyed verbally. Rorty then suggests that we can understand the painting by creating a dialogue with it. In her view, we should *see the painting as a schema for a set of superimposed compositions*, composed of multiple layers: the first layer is the colours, the second is the light and shadow, and then the cut lines of the objects, followed by the objects and the relations between them. Finally the superimposition of these planes should be considered. In contrast to Rorty's focus on guiding observers to understand the message of the painting by creating a dialog with it, Norman [11] first aims to encourage her pupils to rely on their own ability to "read" a piece of art. Only after they have acknowledged their ability to "read" the painting, it is time to teach them specific techniques. Norman's techniques are more about being able to concentrate on one piece of art for a long time, and that itself will bring the observer to understand more about the message the artist intended to convey.

When we are interested in something, we first of all look at it and examine it. Hence, in light of the aforementioned studies and with the purpose of observing differences between professionals and lay observers, we explored the potential of using a mobile eye-tracker for tracking museum visitors' gazing and we present a method to analyse users' behaviour while they look at paintings. The mobile eye-tracker allows us to follow their focus of attention as they observe and explore the artworks.

AVI²CH 2020, September 29, Island of Ischia, Italy

We first focus on related works (Section 2), discussing how eye tracking devices were adopted in art appreciation. Next, the Hecht Museum case study is introduced, and the data composition and the data acquisition protocol are explained (Section 3). In Section 4 we explore emerging issues and challenges and we describe the solutions implemented to analyse the data and generate the heat maps. Finally, in Section 5 we present a brief preliminary analysis, followed by our initial conclusions in Section 6.

2 RELATED WORKS

For most people, vision is the most active of the senses, used in daily life to identify and distinguish objects or persons. Especially in visual art, our eyes are the most useful tool we have for perceiving the messages conveyed by works of art. Considering the importance of this sense, many researchers have investigated the possibility of extracting information from eye-gaze patterns without the support of other subjective tools such as questionnaires and interviews. In recent decades, scholars have focused their attention on eye-tracking devices, and nowadays, mobile eye-trackers are prevalent in many research areas [9].

Understanding user behaviour and processes of visual art appreciation are interesting challenges for which mobile eye-trackers are effective tools. The use of these devices in visual perception evaluation is not new and examples can be found in art [12] and other fields [8]. For a long time, stationary eye-trackers were used in such studies due to their accuracy, despite their high cost. Massaro et al. [7] discuss two known approaches for understanding the process of perception: bottom-up and top-down. The first one is used when the sensory input, the stimulus, gives rise to a data-driven process of decision making about the informative flow. The second one is defined as the development of predictive pattern recognition processes through the use of *a-priori* and contextual information. Usually, top-down processes prevail, or, tend to mask the low-level visually-driven bottom-up processes. In particular, visual exploration of complex images usually follows, in humans, a gazing behaviour strongly influenced by the top-down decision strategy forcing the bottom-up perceptual direction.

Quiroga et al. [12] observed users' behaviour while looking at paintings of various artists and at corresponding modified versions in which different aspects of these art pieces were altered with simple digital manipulations. They discovered common patterns of viewing among the users that create a basic pattern of fixations; for example, the attention of most of the subjects was attracted to the zone in the painting with the sharpest resolution.

In other studies, eye-trackers are used to examine common observation patterns. For example, starting from the idea that Caravaggio consciously constructed a narrative path in his works, Balbi et al. [1] observed and measured the gazing patterns of two samples of volunteers interacting with two Caravaggio paintings in different contexts, in order to test the artist's ability to guide the reader through a visual pathway.

Other studies focus on people's individual cognitive differences, assuming that they influence users' experiences. Raptis et al. [13] proposed a cognition-centered personalisation framework for delivering cultural-heritage activities, tailored to the users' cognitive characteristics. For evaluating it and improving the external validity of the experimental results, they conducted two eye-tracking between-subjects user-studies ($N = 226$) covering two different cognitive styles (field dependence–independence and visualizer–verbalizer) and two different types of cultural activity (visual goal-oriented and visual exploratory). In another work from Raptis et al. [14], eye-tracker are used to study how individual differences in perception and visual information processing affect users' behaviour and immersion in mixed-reality environments.

Following the path of the above studies, in this paper we propose a tool to visually represent gaze data acquired from a mobile eye-tracker, in order to understand visitors' behaviour while looking at paintings.

3 HECHT MUSEUM CASE STUDY

Here we discuss the data available to be processed using the method that will be described in Section 4. The data was acquired in the Hecht Museum in Haifa. Previous research (Kuflik et al. [9]) conducted at the Hecht Museum showed that the art gallery wing of the museum is ideal for the use of a mobile eye-tracker, as it is well suited to the device's field of view both in elevation and left to right angles, given the visitors' height and their standing distance. The Hecht Museum's location on a university campus, where it also serves as a classroom, motivated the use of mobile eye-trackers in art appreciation education. The idea is to examine the gazing patterns of the students of an introductory undergraduate level art history course which aims to develop skills for artwork analysis. The students were asked to observe the same artworks at the beginning and at the end of the semester, before and after acquiring tools in art analysis. Our assumption was that knowledge acquired in class will change the way they look at artworks and these changes could be noticed in the heat maps

3.1 Data description

The data for the study was collected by the Haifa team during a one-semester introductory undergraduate level art history course. The data was collected using the Pupil-Lab[1] Eye-Tracker. As shown in Figure 1, it has two cameras, one which records the eye movements (henceforth "eye camera"), and another which records the environment the user is looking at (henceforth "world camera").

The Haifa research team collected data from thirteen students in two sessions, one before the start of the course and one at the end. During the sessions the students were free to look at a set of predefined artworks as they liked and without time constraints. Therefore, the data is not homogeneous with regard to time, and it is highly variable with regard to how the students observe the artwork. While some users preferred to observe the work of art from close up, trying different observation angles in close examination of the details, others preferred to keep their distance while standing still, examining the artwork from one angle only. This kind of data poses non-trivial challenges. In fact, tracking gazes pointing at a specific stimulus is difficult without using some kind of marker. Typically, the objects to be tracked are identified by physical markers

[1]http://pupil-labs.com/pupil

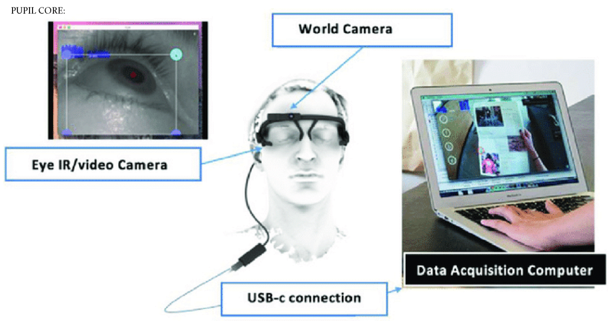


Figure 1: Pupil-Lab Eye-Tracker setting.

positioned on their boundaries, as shown in Figure 2. This solution is not acceptable in the case of art exhibits. First, it is not always possible to place markers on or next to an artwork. Second, visible objects overlaying part of the artwork could compromise the user's experience. Therefore, we addressed this challenge by using computational vision algorithms, to be described in Section 4. The data set contains gaze data about statuettes and paintings, but in this work we focus only on paintings since they can be represented as 2D images. The problem of recognising 3D objects must be addressed differently and is not discussed in this paper. In particular, for initial analysis, the paintings under discussion are *Levona Benschen (Blessing the Moon)* by Jakob Steinhardt and *The Fisherman* by Reuven Rubin (see Figure 3).

We obtained several data files for each participant but we considered only two that contain the information relevant to our goal.The first is the video recorded by the eye-tracker world camera and shows the user's point of view. The second contains time-stamped gazing information for each video frame. In particular, each row has the timestamp, the video frame, a level of confidence, 2D coordinates relative to the video and 3D coordinates relative to the eye-tracker. The way we use these files is discussed in the following sections.

Figure 2: Object tracking using markers. They are positioned in the four corners identifying the area of interest.

4 DATA ANALYSIS

Here we discuss our three-step process of analysing and visualising users' gaze patterns. The steps are:

(a) Reuven Rubin, *The Fisherman*, 1926. Oil on Canvas. Courtesy of the Hecht Museum, University of Haifa, Israel

(b) Jakob Steinhardt, *Levona Benschen*, 1920. Oil on Canvas. Courtesy of the Hecht Museum, University of Haifa, Israel

Figure 3: Paintings considered to test the tool.

(1) Input data manipulation and filtering
(2) Data elaboration and reference system switching
(3) Heat map generation

For each phase a pertinent python module was developed. Each module is independent, making it possible to reproduce only a single phase if necessary. Moreover, the modularity grants us better software maintainability. In the next sections we provide an in-depth explanation of each phase.

(a) Original video light conditions.

(b) The same frame after gamma and saturation were modified.

Figure 4: Showing the same video frame before (4(a)) and after (4(b)) modifying gamma and saturation.

4.1 Input data manipulation and filtering

The first phase consists of manipulating and filtering the data provided by the eye-tracker. It requires three input files: the video recorded by the world camera, the csv file containing the gaze, data, and a good quality image of the painting, from now on referred to as the reference image, to be matched in the video. During this phase, as the csv file is not suitable for the next elaboration step, we need to manipulate it by removing useless columns and switching the file format to tsv.

Moreover, we noticed that the video light conditions were poor, causing a low matching rate in the second phase. In order to solve this problem, we processed the video by modifying gamma and saturation, making it brighter and improving the performance of next phase. To this end we used the multi-platform tool FFmpeg [2] which provides several video manipulation functionalities. Figure 4 shows a comparison between an original and modified video frame. In the original version (Figure 4(a)), we can see that the shapes in the unlit areas of the painting are not well defined, in contrast to the modified version (Figure 4(b)). Obtained data will be used as input for the second phase.

4.2 Data elaboration and reference system switching

Data manipulated in the previous step is used as the input to the feature matching process described next. We first need to consider the nature of our input files. We have a tsv file where coordinates are expressed in pixels relative to the world camera video. This means that the file contains more information than needed. In fact, we only care about gaze coordinates falling within the painting area. As people are left free to move during the recording session, it is impossible to define the painting area in advance, but the painting must be recognisable in each video frame. Thus the unneeded information must be pruned from the gaze data. Moreover, coordinates need to be translated from the video reference system to the reference image system.

Figure 5: Key-points, represented by the green circles, found using the SIFT [4] algorithm on the reference image.

Mobile Gaze Mapping [5] is the solution we choose to address the above-mentioned problems. Since gaze position is recorded without any reference to fixed objects in the environment, this poses a challenge for studying how an individual views a particular stimulus over time. This toolkit addresses this challenge by automatically identifying the target stimulus in every frame of the recording and mapping the gaze positions to a fixed representation of the stimulus. It does this by identifying matching key-points between the reference stimulus and each frame of the video. Key-points (see Figures 5, 6, 7) are obtained using the Scale Invariant Feature Transform algorithm (SIFT [4]), and, importantly, they are invariant to image feature scale and rotation. Matches between key-points are found using the Fast Approximate Nearest Neighbour search algorithm (FLANN [10]). Both algorithms were implemented in OpenCV [2]. Once matching key-points are identified, we determine the 2D linear transformation that maps key-points from the video frame to key-points on the reference image (see Figure 8). Once determined, this same transformation is applied to the gaze position sample corresponding to the given video frame, resulting in the gaze position being expressed in terms of the pixel coordinate system of the reference image. The transformed image is saved and used in the next phase.

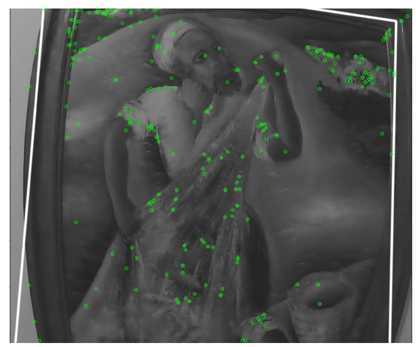


Figure 6: Key-points, represented by the green circles, found using the SIFT [4] algorithm on a world camera frame.

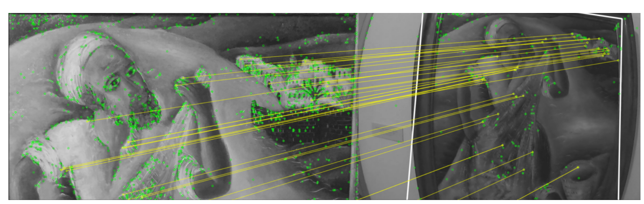


Figure 7: Key-point matching between the reference image and the world camera frame.

The tool was tested for accuracy and precision by MacInnes et al. [6] across three different eye-tracking devices, including the one we used in this work and achieved 97,1% valid gaze points.

[2]https://www.ffmpeg.org/

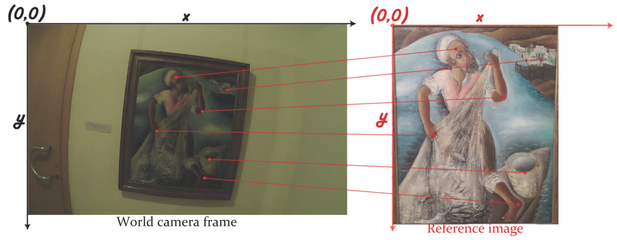
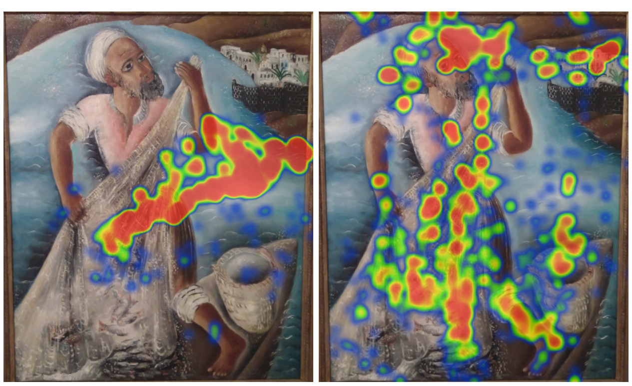


Figure 8: Finding the 2D linear transformation that maps key-points from the video frame to key-points on the reference image.

4.3 Heat map generation

In the third step, a heat map is generated, starting from the transformed gaze data provided by the previous one. Heat maps are an effective data visualisation technique that provide a graphic representation of gaze focus on the painting. By comparing the heat map of the session recorded before the course with the one recorded after, we can interpret the data in a simple manner. We use the Python module heatmappy [3] for this purpose. Heatmappy provides a set of functions to generate and customise heat maps starting from coordinates in the form of tuples. We extract coordinates (x,y) from transformed gaze data and express them in the form required by heatmappy. Now we are able to give the generated tuple as input to the module and generate the heat map. Examples of the resulting images are shown in Figures 9 and 10.

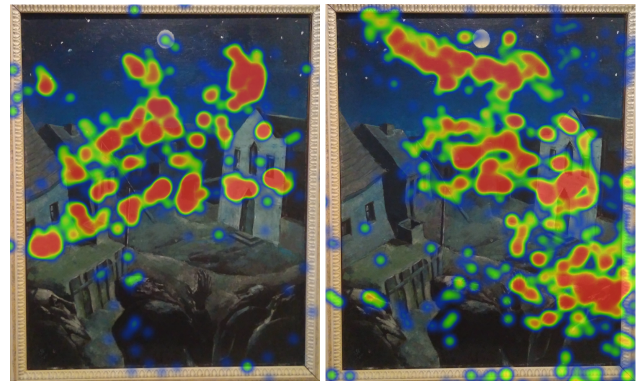


(a) Before. (b) After.

Figure 9: Heat maps of the two sessions generated while Subject 3 observed *The Fisherman.*

5 PRELIMINARY ANALYSIS

Here we discuss the heat maps that were obtained from the data set using our tool. Although ideally the heat maps should be analysed by art experts, here we try to interpret them numerically. Even a quick glance at the examples in Figures 11 and 12 shows noticeable differences between the "before" and the "after". We assumed that these differences could be analysed by examining three metrics:



(a) Before. (b) After.

Figure 10: Heat maps of the two sessions generated while Subject 5 observed *Blessing the Moon* painting.

gaze distribution, gaze pathway and gaze focus on salient parts of the painting. Gaze distribution determines whether the user's gaze is focused on a few points or distributed over many points, and it shows how this behaviour changes between the sessions. Gaze pathway is used to identify common patterns between the users in the pathways of the second recording session. We assume that the gaze pathway might change between sessions. Finally gaze focus helps us to understand whether the user's gaze focuses on the salient parts of the painting as defined by art experts.

In this initial work we focused on the first metric - the gaze distribution. We assume that non-expert users focus their gaze only on a few points having a strong visual impact, while expert users are able to discern more details than non-experts. To estimate the gaze distribution we used the entropy measure. When it is applied to heat maps, it returns a higher value when there are many focal points and a lower value when there are few focal points. The results for all the subjects are presented in Figures 11 and 12. The y-axis shows the subjects participating in the experiment (S1,S2,...,S13) while the x-axis represents the entropy values. For each subject a comparison between the entropy found before the course (blue bar) and after the course (orange bar) is shown. Note that for *The Fisherman* we excluded three recordings because most of time the mobile eye-tracker did not recognise the gaze, probably because of a calibration error. As already noted, Figures 9 and 10 present the heat maps generated while Subject 3 observed *The Fisherman* and Subject 5 observed *Blessing the Moon*. It can be seen that entropy indeed varies together with the gaze distribution. In particular, when the gaze distribution is low, we see contiguous and well-defined areas, but conversely, when the gaze is highly distributed, we see many and less-contiguous areas (see Figure 9(b)).

Entropy levels, in most of the cases, are higher in the session recorded after the course. For *The Fisherman*, the average entropy values are 4.31 (before) and 5.0 (after), and for *Blessing the Moon* they are 5.0 (before) and 5.55 (after), matching our assumption. Also interesting is overall gazing duration. Considering that the time parameter is important for the analysis, we discovered that, in most of the cases, users spent more time looking at the artworks

[3] https://github.com/LumenResearch/heatmappy

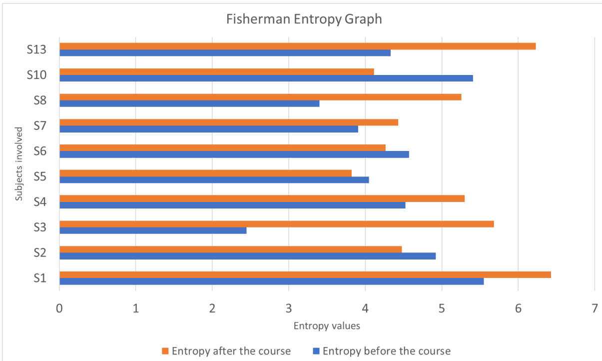


Figure 11: Comparison of entropy levels of the two recording sessions for *The Fisherman* painting.

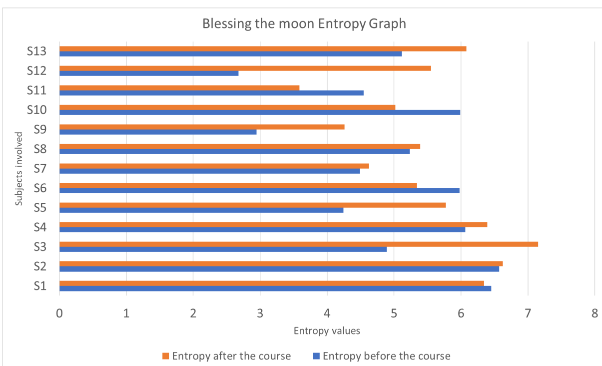


Figure 12: Comparison of entropy levels of the two recording sessions for *Blessing the Moon* painting.

in the second session than the first one. This may be considered as a sign of competence. While during the first session users probably looked at the artworks in a superficial manner, in the second one, after acquiring guidelines for art appreciation, they looked at the artworks more deeply, trying to discern the details. For *Blessing the Moon*, the average observation times for all participants were 30.61s before the course and 35.84s afterwards, while for *The Fisherman*, the average times were 27.4s before the course and 35.1s afterwards. Since this is a work in progress, we just show a preliminary and quite simple analysis based on entropy and observation time, but in the near future we plan to extend the study, together with art experts who will visually examine the heat maps visually in order to validate the current results, and we also plan to develop new analysis tools. Moreover, since in this study we did not consider the eventual influence of the first observation session on the second one, we planned to improve the analysis by involving a higher number of users and splitting them into two groups, one who will see the paintings before and after the semester, and one who will see the paintings only after the semester. By doing that, we will be able to make a statistical analysis of the two groups, evaluating the entity of the bias introduced by the first observation.

6 CONCLUSIONS

Understanding how people observe works of art, in our specific case paintings, is an open challenge and mobile eye-trackers are becoming the principal technology to address it. However, manual analysis of the data is impractical. For gaze_positions.csv for example, we had about 5000 rows for 43 seconds video. Therefore, an automatic tool that minimises the time spent on manual processing and exposes the data in a simple and intuitive format, could facilitate further studies in this field by allowing researchers to focus their effort on data interpretation. In this work we presented and implemented a pipeline to automatically filter, process and present mobile eye-tracker data. The data is presented through heat maps that can be easily interpreted by art experts. Moreover, we proposed a preliminary numerical analysis based on a calculation of heat map entropy calculation. In any case this is a preliminary work that poses new challenges to be addressed. In future works we aim to expand the tool's features, introducing new metrics suggested by experts, and providing the option to extract and manage data from eye-tracking data recordings of 3D objects.

Heat maps and metrics provided by the tool could be used by museum curators and teachers to improve their work. For example, a teacher could understand if his students are acquiring the skills he explain and, if they are not, he could change the teaching approach. A museum curator could use the tool to understand if the painting position and the light condition emphasise the areas he wanted to. Finding in the heat maps that the the gaze rarely focuses on these areas could suggest that something needs to be changed in the artwork exposition.

7 ACKNOWLEDGEMENTS

The authors thank Menachem Merzel from the Information Systems Department at the University of Haifa for the data he collected, the students from the artwork analysis course of the Art History department who volunteered and contributed to the study, and the Hecht Museum for allowing us to perform the study. This work is funded by the Italian PRIN project *Cultural Heritage Resources Orienting Multimodal Experience* (CHROME) #B52F15000450001.

REFERENCES

[1] Barbara Balbi, Federica Protti, and Roberto Montanari. 2016. Driven by Caravaggio Through His Painting. An Eye-Tracking Study. *COGNITIVE 2016, The Eighth International Conference on Advanced Cognitive Technologies and Applications*, 72–26.

[2] Gary Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* 25 (01 2000).

[3] M. Housen. 1999. Eye of the Beholder : Research, Theory, and Practice. *Aesthetic and Art Education : a Transdisciplinary Approach* (1999). https://ci.nii.ac.jp/naid/10020820591/en/

[4] David Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60 (11 2004), 91–. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[5] Jeff MacInnes, Shariq Iqbal, John Pearson, and Elizabeth Johnson. 2018. Mobile Gaze Mapping: A Python Package for Mapping Mobile Gaze Data to a Fixed Target Stimulus. *Journal of Open Source Software* 3, 31 (2018), 984. https://doi.org/10.21105/joss.00984

[6] Jeff MacInnes, Shariq Iqbal, John Pearson, and Elizabeth Johnson. 2018. Wearable Eye-tracking for Research: Automated Dynamic Gaze Mapping and Accuracy/Precision Comparisons across Devices. https://doi.org/10.1101/299925

[7] Davide Massaro, Federica Savazzi, Cinzia Di Dio, David Freedberg, Vittorio Gallese, Gabriella Gilli, and Antonella Marchetti. 2012. When Art Moves the Eyes: A Behavioral and Eye-Tracking Study. *PLOS ONE* 7, 5 (05 2012), 1–16. https://doi.org/10.1371/journal.pone.0037285

[8] S. Mateeff. 1980. Visual Perception of Movement Patterns during Smooth Eye Tracking. *ACTA PHYSIOL. PHARMACOL. BULG.; ISSN 0323-9950; BGR; DA. 1980; VOL. 6; NO 3; PP. 82-89; ABS. RUS; BIBL. 13 REF.* (1980).

[9] Moayad Mokatren, Tsvi Kuflik, and Ilan Shimshoni. 2017. Exploring the Potential of a Mobile Eye Tracker as an Intuitive Indoor Pointing Device: A Case Study in Cultural Heritage. *Future Generation Computer Systems* (07 2017). https://doi.org/10.1016/j.future.2017.07.007

[10] Marius Muja and David Lowe. 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *VISAPP 2009 - Proceedings of the 4th International Conference on Computer Vision Theory and Applications* 1, 331–340.

[11] Jane Norman. 1970. How to Look at Art. *The Metropolitan Museum of Art Bulletin* 28, 5 (1970), 191–201. http://www.jstor.org/stable/3258480

[12] Rodrigo Quian Quiroga and Carlos Pedreira. 2011. How Do We See Art: An Eye-Tracker Study. *Frontiers in Human Neuroscience* 5 (2011), 98. https://doi.org/10.3389/fnhum.2011.00098

[13] George Raptis, Christos Fidas, Christina Katsini, and Nikolaos Avouris. 2019. A cognition-centered personalization framework for cultural-heritage content. *User Modeling and User-Adapted Interaction* 29 (02 2019). https://doi.org/10.1007/s11257-019-09226-7

[14] George E. Raptis, Christos Fidas, and Nikolaos Avouris. 2018. Effects of mixed-reality on players' behaviour and immersion in a cultural tourism game: A cognitive processing perspective. *International Journal of Human-Computer Studies* 114 (2018), 69 – 79. https://doi.org/10.1016/j.ijhcs.2018.02.003 Advanced User Interfaces for Cultural Heritage.

[15] Amelie Rorty. 2014. Dialogues with Paintings: Notes on How to Look and See. *The Journal of Aesthetic Education* 48, 1 (2014), 1–9. https://www.jstor.org/stable/10.5406/jaesteduc.48.1.0001

[16] Bruno Zevi, Milton Gendel, and Joseph A. Barry. 1957. Architecture as Space. How to Look at Architecture. *Journal of Aesthetics and Art Criticism* 16, 2 (1957), 283–283. https://doi.org/10.2307/427623