The Art of Pervasive Eye Tracking
Unconstrained Eye Tracking in the Austrian Gallery Belvedere

Thiago Santini
University of Tübingen
Tübingen, Germany
thiago.santini@uni-tuebingen.de

Hanna Brinkmann
University of Vienna
Vienna, Austria
hanna.brinkmann@univie.ac.at

Luise Reitstätter
University of Vienna
Vienna, Austria
luise.reitstaetter@univie.ac.at

Helmut Leder
University of Vienna
Vienna, Austria
helmut.leder@univie.ac.at

Raphael Rosenberg
University of Vienna
Vienna, Austria
raphael.rosenberg@univie.ac.at

Wolfgang Rosenstiel
University of Tübingen
Tübingen, Germany
wolfgang.rosenstiel@uni-tuebingen.de

Enkelejda Kasneci
University of Tübingen
Tübingen, Germany
enkelejda.kasneci@uni-tuebingen.de

Figure 1: Eye-tracking enabled insights. Whereas an external observer might be tempted to consider that the visitor is gazing at the face from Gustav Klimt’s portrait of “Amalie Zuckerandl” because of its saliency, eye tracking reveals the visitor’s true fixation position, suggesting an analysis of the painting details along Amalie’s dress. Figure best visualized in digital form.

ABSTRACT
Pervasive mobile eye tracking provides a rich data source to investigate human natural behavior, providing a high degree of ecological validity in natural environments. However, challenges and limitations intrinsic to unconstrained mobile eye tracking makes its development and usage to some extent an art. Nonetheless, researchers are pushing the boundaries of this technology to help assess museum visitors’ attention not only between the exhibited works, but also within particular pieces, providing significantly more detailed insights than traditional timing-and-tracking or external observer approaches. In this paper, we present in detail the eye tracking system developed for a large scale fully-unconstrained study in the Austrian Gallery Belvedere, providing useful information for eye-tracking system designers. Furthermore, the study is described, and we report on usability and real-time performance metrics. Our results suggest that, although the system is comfortable enough, further eye tracker improvements are necessary to make it less conspicuous. Additionally, real-time accuracy already suffices for simple applications such as audio guides for the majority of users even in the absence of eye-tracker slippage compensation.

CCS CONCEPTS
• Human-centered computing → Ubiquitous and mobile computing systems and tools; Empirical studies in ubiquitous and mobile computing; Empirical studies in HCI; • Computer
systems organization → Embedded systems; • Applied computing → Fine arts;

KEYWORDS
Eye tracking, pervasive, system, mobile, embedded, real-time, pupil detection, pupil tracking, gaze estimation, calibration

ACM Reference Format:

1 INTRODUCTION
Pervasive mobile eye tracking provides a rich data source to investigate human natural behavior [Kasneci 2017; Wessel et al. 2007], which provides a higher degree of ecological validity than traditional in-the-lab controlled experiments, specially when combined with natural environments. In the context of museums, there is strong evidence that ecologically valid testing in natural conditions (e.g., in museum conditions) is paramount for experimental aesthetics [Walker et al. 2017], and object authenticity (e.g., photographs vs real objects) has a positive impact on subjects’ attention [Filippini Fantoni et al. 2013]. Moreover, eye tracking provides significantly more detailed insights than traditional timing-and-tracking or external observer approaches [Filippini Fantoni et al. 2013] — e.g., see Fig. 1 and Fig. 2. Previous works using mobile eye trackers in museums have investigated the difference in eye movements between children and adults [Walker et al. 2017] and how expertise influences viewing behavior of domestic textiles [Tatler et al. 2016], as well as performed exploratory studies [Filippini Fantoni et al. 2013; Mayr et al. 2009; Mokatren et al. 2016; Wessel et al. 2007]. However, many of these works are either constrained (e.g., subjects observe paintings only from a fixed position), have very short durations, or do not discuss the accuracy of the eye trackers, which is known to still suffer from multiple issues such as drift and parallax errors [Holmqvist et al. 2011]. In this paper, we present in detail the eye tracking system used in a large scale fully-unconstrained study in the Austrian Gallery Belvedere, providing useful information for system designers. This study is described in detail in Section Method and is part of a larger experiment, whose aim is to investigate how visitors perceive artworks in relation to distinct museological settings1. Furthermore, we report on usability and real-time performance metrics.

2 EYE-TRACKING SYSTEM
When designing the eye-tracking system, one of our goals was to make the system as general as possible while maintaining real-time and pervasive in-device eye tracking functionality. This includes being accessible for users with glasses, which represent about 30 percent of the young adult population [Morgan and Rose 2005], without the need for them to remove their glasses — e.g., as required by Tobii and SMI glasses. A direct result from these requirements is that the resulting system can be easily calibrated and used in real-time in an individual fashion, enabling future gaze-based human-computer interaction applications [Bulling and Gellersen 2010] — e.g., gaze-activated audio guides. Research wise, the developed system also allows an experimenter to check pupil detection and gaze estimation accuracy in real time without the need of additional devices and available network, allowing for significant improvement in data quality as well as a much simpler and cheaper experimental setup. In comparison, other mobile eye-tracking systems commonly require a separate recording unit and a separate control device for the experimenter — e.g., Dikablis, SMI, and Tobii glasses [Ergoneers 2018; SensoMotoric Instruments 2018; Tobii 2018].

2.1 Hardware
We employed a binocular Pupil Labs [Pupil Labs 2018] head-mounted eye tracker (2x Pupil Cam2 eye cameras; 1x Pupil Cam1 scene camera). This device was paired with a Microsoft Surface Pro 42

1Combinations of artworks, spacial placement, additional textual information, etc.

2Configured with an Intel® Core™ i5-6300U

Figure 2: Eye-tracking enabled insights. Eye tracking reveals the trajectory of fixations (i.e., the scanpath) as the visitor attends to Giovanni Segantini’s The Evil Mothers. These scanpaths provide key insights into the human cognitive process and are also useful for distinguishing a subject’s expertise level [Kübler et al. 2017]. Figure best visualized in digital form.
provided ideal places to run the eye tracking cable and place the remote-control receiver. The complete system (see Fig. 3) weighs less than 1 kg with a total cost of about 3197 EUR – contributed mostly by the eye tracker (2150 EUR) and tablet (1015 EUR); user feedback is reported in Section Usability. Thanks to this small cost, we were able to conduct the experiment with four systems simultaneously; in contrast, solutions provided by vendors typically cost more than our four systems combined. An additional benefit enabled by multiple systems is the ability of running simultaneous measurements within groups, allowing for the exploration of collaborative learning and social engagement [Mayr et al. 2009]. It is worth noting that this social aspect adds to the ecological validity of museum-related experiments as approximately 80 to 95 percent of visits are done in groups [Hein 2002].

2.2 Software

Although the Pupil Labs eye tracker manufacturer provides a software solution (Pupil Capture and Pupil Player), we opted to use EyeRecToo [Santini et al. 2017b] instead. One of the main reasons for this decision was the difficulty in calibrating outside of the lab in natural environments with Pupil Capture and its manual marker calibration. Additionally, the pupil detection we experienced seems to be bimodal, either working relatively well or not at all. EyeRecToo on the other hand provides a robust marker detection (through ArUco markers [Garrido-Jurado et al. 2014]) and an advanced calibration method called CalibMe [Santini et al. 2017a], which uses a target moving w.r.t. the scene camera for calibration (see Section Method). Furthermore, EyeRecToo also provides PuRe [Santini et al. 2018a], a more robust pupil detector; we also have extended this pupil detector with a pupil tracking algorithm that significantly outperforms existing pupil detection methods. Although this tracking algorithm is not a contribution of this paper, we provide performance figures (see Fig. 4) to contextualize system performance based on five available challenging and realistic data sets provided in previous work, namely, the Świrski [Świrski et al. 2012], ExCùSe [Fuhl et al. 2015], ElSe [Fuhl et al. 2016b], LPW [Tonsen et al. 2016], and PupilNet [Fuhl et al. 2016a] data sets. This novel tracking algorithm is described in detail in [Santini et al. 2018b].

2.3 Advice for System Designers

Pervasive robust video-based eye tracking remains not only challenging, but also computationally expensive. Since embedded devices are required to allow for mobility, handling three camera streams, image processing, and data recording is not trivial. In particular, one of the biggest challenges we found was performance throttling mechanisms – e.g., to stay within thermal constraints [Rao and Vrudhula 2007]. Such mechanisms are present virtually in every powerful modern system on chip. For instance, the Intel Extreme Tuning Utility [Intel 2018] reports four different throttling mechanisms for the device employed in this work: Thermal, power limit, current limit, and motherboard VR thermal throttling. While some of these mechanisms are activated to keep the device

---

1In our study, 70.64 percent of visitors reported being accompanied during their visit. Pupil Labs seems to be aware of this issue since modifications to the marker were made recently; see: https://github.com/pupil-labs/pupil/releases/tag/v1.2

5The actual number of cameras might change depending on the eye tracker. For instance, a monocular one has two cameras, whereas Tobii glasses have five.
Despite the improved pupil detection rates, some subjects’ pupils were still hard to track some due to occlusions from eyelids and glasses.

Within thermal constraints or because the battery can no longer supply enough current, the end result is more or less the same: A significant performance drop. Practitioners should be aware of such behavior and test during longer periods of operation to assure that performance requirements are within the expected range. Failing to do so might result in severe frame dropping during field experiments; it is worth noting that EyeRecToo provides a useful performance monitoring widget, which allows one to monitor frame dropping. Furthermore, when developing mobile systems such as ours, it is important to minimize single points of hardware failure, such as USB connections. During our initial in-the-lab tests, we noticed that the connection between the USB hub and the eye tracker cable would sparsely stop working because of cable movement; a simple solution with Velcro was promptly arranged to make sure the connection remained tight even during harsh movements. In contrast, we experienced connection issues with the Pupil Labs USB clip only a single time, during the field experiment. Finally, it is also important to be aware of operating system specific requirements and caveats. For Windows 10.1, we found the following worth of mention: 1) disabling the Connected Standby feature, which allows for a more extensive customization of advanced power settings, 2) disable sleeping/hibernation options, 3) disable USB selective suspend setting, 4) disable Turn off hard disk after, 5) uninstall Microsoft OneDrive, 6) disable the Windows search index service, 7) disable the Windows updater service after every reboot. The latter three items are necessary to prevent Windows from running CPU intensive tasks during experiments.

3 METHOD
3.1 Participants
Visitors arriving at the top of the Grand Staircase of the Austrian Gallery Belvedere were invited to join the experiment. The only requirements for a participant to take part was their consent (thus requiring them to be 18 or older) and ability to speak either English or German. The experiment took place from the 22nd to 28th of January 2018, with data being collected during five of these days. In total 109 subjects (63 females) took part in our experiment averaging 34.86 years of age ($\sigma = 14.62$).

3.2 Procedure and Stimuli
Upon accepting to take part in the experiment, the subject received and read a consent form containing experiment instructions. Afterwards, an experimenter assisted the subject to don the eye-tracking system (see Section Eye-Tracking System for details). The subject was then asked to stand on a floor mark approximately 1.16 m away from a CalibMe collection marker that stood at about the same height as the paintings in the gallery. The experimenter adjusted the scene camera to center the collection marker, and then proceeded to adjust the eye cameras to a suitable position. Subsequently, the user was instructed on how to perform the calibration by always gazing at the central intersection of the collection marker while moving his head in a spiral fashion smoothly and slowly. The experimenter then started the recording and the subject calibration procedure, controlling when the eye tracker calibration started and stopped. After calibration, the subject was asked to gaze at four Post-its® about 25° away from the marker center so the experimenter could check the gaze estimation quality. It is worth noting that the experiment was conducted by ten experimenters without previous experience with head-mounted eye trackers, which prior to the experiment had a half hour instruction session with an eye-tracker expert with about three years of experience. The participant was then instructed to freely roam through four rooms (containing more than 30 distinct paintings and sculptures) as he/she wished, and that the experimenter would meet him at the end of the last room. At the end of the visit, the experimenter and subject proceeded to a separate room where: 1) a second calibration was performed, 2) the subject was interviewed and performed a remembrance mapping task, and 3) the subject answered a questionnaire containing please note that most of these informations came from extensive interactions with Microsoft Support. These changes are provided as part of a detailed documentation of our system without warranty of any kind. In no event shall the authors be liable for any claim, damages or other liability.

<table>
<thead>
<tr>
<th>Pixel Error</th>
<th>Tracking (Ours)</th>
<th>PuRe</th>
<th>Pupil Labs 2D</th>
<th>Pupil Labs 3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>0.9</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: On the top, the cumulative detection rate for the aggregated 266,786 images from all data sets. On the bottom, run time distribution across all evaluation images using a Tukey schematic boxplot [Frigge et al. 1989]; results without outliers and based on a Intel® Core™ i5-4590 CPU. Despite the improved pupil detection rates, some subjects’ pupils were still hard to track some due to occlusions from eyelids and glasses.
museum-visit and eye-tracking related questions. After the experiment, participants were rewarded with a small souvenir.

4 RESULTS

In this section, we report initial results obtained from questionnaires and from real-time system performance – i.e., as obtained during the experiment without post processing. Post processing of the data (e.g., improving pupil detection for cases in which pupil edges are not visible, compensating for eye tracker slippage) is currently under way and left for future work.

4.1 Usability

For the usability analysis, we have included all participants. Since the museum-visit-related questionnaire and interview already lasted a considerable amount of time, we opted to include only seven eye-tracking-related questions as to not overload subjects. Participants were asked “Please indicate how much you agree with the following statement” for each of the statements listed in Fig. 5. This figure also shows the distribution of the subjects’ feedback. Q1 addresses one of the main issues with eye tracking: The calibration, which has been considered one of the main factors hindering eye tracking adoption [Morimoto and Mimica 2005] and is often considered of poor usability, experienced as difficult, and described as tedious [Pfeuffer et al. 2013]. In this regard, CalibMe was perceived by ≈92% of users as easy to follow, even when instructed by non-experienced experimenters. Q2, Q3, and Q4 address comfort and obtrusiveness [Mayr et al. 2009; Poole and Ball 2006]. Feedback from these questions show that 1) the small-footprint 3D-printed eye tracker produced by Pupil Labs has a good comfort rating (agreed by > 84% participants) without limiting visibility – agreed by > 86% of participants, and 2) the whole system is perceived as comfortable while maintaining mobility, real-time gaze estimation capability, and good autonomy (≈ one and a half hour of recording and processing). Q5 and Q6 probe the implications of head-mounted eye tracking for the ecological validity of data collected with such devices: Even with the unobtrusiveness of the system, 31.5% of participants feeling moderately observed and 6.48% observed. Nonetheless, 76.1% of the participants felt that the eye tracking equipment did not interfere with their art perception. Additionally, we also inquired the subjects at which point in time they forgot about the eye-tracking equipment, with the majority reporting they forgot it within two minutes into the experiment as shown in Fig. 6. Nevertheless, a significant part (33.94%) reported that they never forgot about the device, possibly because the head-mounted eye tracker used still is too salient in the subjects’ field of view. This suggests that further improvements are still needed to make eye trackers less conspicuous.

4.2 Real-Time Performance and Metrics

To estimate initial figures for the real-time performance, we excluded thirteen out of 109 participants for different reasons: a) five subjects manually moved the eye tracker significantly, b) three subjects went to the wrong section of the museum, c) for two an

Figure 5: Distribution of participant’s answers to the eye-tracking usability questionnaire. For the top two questions (in green), the more the subjects agree, the better. For the remaining four questions (in red), the more they disagree, the better.
alternative calibration method was experimented with, d) for one the USB clip disconnected, e) one exhibited nystagmus during calibration, and f) for one an experimenter forgot to start the recording.

The calibration at the beginning of the experiment had an average duration of 23.25 seconds ($\sigma = 10.08$), during which 438.57 relationships between pupil and gaze positions were collected on average ($\text{sigma} = 197.13$). From these, on average 35.50 ($\sigma = 35.06$) were removed as outliers automatically – see [Santini et al. 2017a] for outlier identification details. Fig. 7 shows the distribution of these points w.r.t. the scene camera image as a 2D histogram smoothed with a 5 px Gaussian to improve visualization.

Several eye-tracking studies will only shortly report on the accuracy of the gaze signal by reiterating a manufacturer provided number – usually below one degree – that does not correspond to the actual accuracy during real usage. For instance, real accuracies reported are as large as 2° [Holmqvist et al. 2011] for tower-mounted eye trackers. Head-mounted devices tend to exhibit even worse accuracies due to eye-tracker slippage [Kolakowski and Pelz 2006], oculomotor jitter [Martinez-Conde et al. 2004], parallax error [Mar-danbegi and Hansen 2012], and even changes in pupil dilation that can result in inaccuracies up to 1.5° [Holmqvist et al. 2011]. In fact, some works report average errors of 3.6° even for high-grade; eye trackers with multiple glint sources and reflection filters [Schüssel et al. 2016], whereas others consider accuracies within 5° as acceptable for single-glint eye trackers [Kübler et al. 2015].

Figure 6: Distribution of counts regarding at which point in time participants forgot about the device during the experiment.

Figure 7: Distribution of calibration points w.r.t. to the scene camera image. There is a clear bias on the center because the camera was adjusted for the marker to be center. Nonetheless, notice how the points are well spread around the scene camera image, providing large interpolation areas for the gaze estimation.

5 FINAL REMARKS

In this paper, we presented in detail the eye tracking system used in a large scale fully-unconstrained study in the Austrian Gallery Belvedere. Our usability results show that while the calibration and comfort of the system is already at an acceptable level, further improvements are necessary to make head-mounted eye trackers more inconspicuous. From an accuracy point of view, while about 20% of subjects had an accuracy within 5°, whereas 43.33% still retained such accuracy at the end despite lack of eye-tracker slippage compensation and in such an unconstrained scenario. Please notice that we approximated the error in visual angle based on the pixel error using a linear relationship and the scene camera’s field of view. Since head movements become a regular feature of gaze shifts at about 20° of visual angle [Becker 1989], the point of regard usually falls within the central region of the camera image; thus, this approximation overestimates the error due to the camera’s fish eye lenses. For simple applications in which lower accuracies are acceptable, such as gaze-based audio guides, the employed system already suffices for the majority of subjects even during longer visits.
Currently, future work focuses on two main points. First, to investigate solutions to eye-tracker slippage – e.g., by employing eye models to determine the relative pose of the eye to the its respective camera, and estimating the position of the eye cameras to the field camera during calibration through CalibMe’s collection marker pose. Second, how to automatically produce semantic mapping from the scene camera images – i.e., to give a semantic meaning to the image region surrounding the gaze estimate. We are currently investigating approaches to achieve this mapping. For instance, by employing keypoints (such as ORB [Rublee et al. 2011] and BRIEF [Calonder et al. 2010]) or by using convolutional neural networks (CNNs) for segmentation (such as Mask R-CNN [He et al. 2017]) to determine each painting/statue’s pose and identification, allowing us to map fixations relative to the scene camera to objects of interest. Similarly, by using state-of-the-art CNNs for face detection [Jiang and Learned-Miller 2017] and recognition [Parkhi et al. 2015], it is also possible to estimate social interaction, an approach that can be easily extended to other social environments – e.g., classrooms [Santini et al. 2017c]. Furthermore, the object-of-interest’ pose can also provide meaningful cues such as interaction distance range. Although we did not measure participant willingness to wear the eye-tracking system, anecdotally we did not experience negative responses from the visitors towards the system. Nonetheless, this is a further usability question that should be considered in the future. Moreover, since traditional gaze-based interaction techniques (e.g., dwell time [Duchowski 2002]) might prove insufficient for such dynamic scenarios, multimodal interfaces should be considered. Finally, it is worth considering remote gaze-sensing solutions, such as EyePhlance [Shell et al. 2003], which might offer a less intrusive solution to map visitors’ attention.

REFERENCES

Figure 8: On the left, cumulative gaze-estimation error distribution for the auto-evaluation points from the initial calibration (Start) and from collected points at the last calibration (End). Results are shown for each participant’s third quartiles ($Q_3$) to exclude possible outliers such as blinks and gaze distractions. On the right, error range overlayed on top of a scene image extracted from our experiment for reference; the circles (radius) represent approximately 1°, 3°, 5°, and 9°. Figure best visualized in digital form.