Julian Steil, Marc Tonsen* *Max Planck Institute for Informatics, Saarland Informatics Campus, Germany*
Yusuke Sugano* *Institute of Industrial Science, The University of Tokyo, Japan*
Andreas Bulling* *University of Stuttgart, Institute for Visualization and Interactive Systems, Germany*
**Work conducted while at the Max Planck Institute for Informatics*

Editors: Nic Lane and Xi Zhou

InvisibleEye:
Fully Embedded Mobile Eye Tracking Using Appearance-Based Gaze Estimation

Excerpted from "InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation" from *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* with permission. https://dl.acm.org/citation.cfm?id=3139486.3130971

Despite their potential for a range of exciting new applications, mobile eye trackers suffer from several fundamental usability problems. *InvisibleEye* is an innovative approach for mobile eye tracking that uses millimeter-size RGB cameras, which can be fully embedded into normal glasses frames, as well as appearance-based gaze estimation to directly estimate gaze from the eye images. Through evaluation on three large-scale, increasingly realistic datasets, we show that InvisibleEye can achieve a person-specific gaze estimation accuracy of up to 2.04° using three camera pairs with a resolution of only 3x3 pixels.

Human gaze has a long history as a means for fast, accurate, and natural interaction with both ambient [15] and body-worn displays, including smartwatches [3] and has, more recently, also been shown to be a rich source of information about the user [6]. Eye movements are closely linked to everyday human behavior and cognition and can therefore be used for computational user modeling, such as for eye-based recognition of daily activities [2], visual search targets [8], or personality traits [4] – including analyses over long periods of time for life-logging applications [9]. Interest in gaze has been fuelled by recent technical advances and significant cost reductions of mobile eye trackers that can be worn in daily life and provide insights into users' everyday gaze behavior [1].

Despite its appeal, mobile eye tracking suffers from several fundamental usability problems. First, current mobile eye trackers are still uncomfortable to wear, especially over long time periods: The required high-quality imaging sensors are large and thus often occlude the user's field of view, are heavy and cause discomfort or even pain. Second, current eye trackers limit users' mobility given that they require a wired connection to a recording computer. Finally, their obtrusive design leads to low social acceptance and unnatural behavior of both the wearer and people they interact with [7], thus fundamentally limiting the practical usefulness of mobile eye tracking as a tool in the social and behavioral sciences.

We argue that it is ultimately necessary to fully integrate eye tracking into regular glasses, i.e., to effectively make eye tracking visually and physically unnoticeable to both the wearer and bystanders. A key requirement for such unnoticeable (*invisible*) integration is to reduce the size of an eye tracker's core component: the imaging sensors. Smaller sensors not only significantly reduce the device's weight, but can also be positioned in the visual periphery to avoid occlusions within the users' field of view. In addition, the low resolution common to these sensors generates significantly less data that could be processed on the device itself, stored, or transmitted wirelessly, thus removing the need for a separate recording device. Finally, the reduced computation required to process low-resolution images decreases the load on the processor and, as such, helps to extend the recording time beyond the current limit of only a few hours.

The eye tracker we developed, InvisibleEye, can be fully embedded into a normal glasses frame (see Figure 1, bottom left). To achieve this, we took a radically different approach to mobile eye tracking. Instead of a single camera and model-based gaze estimation, InvisibleEye uses multiple, millimeter-size imaging sensors positioned around the eye as well as a computational method based on an artificial neural network for so-called appearance-based gaze estimation – estimating gaze of a specific user by automatically analyzing the eye images obtained from the cameras. Here we briefly

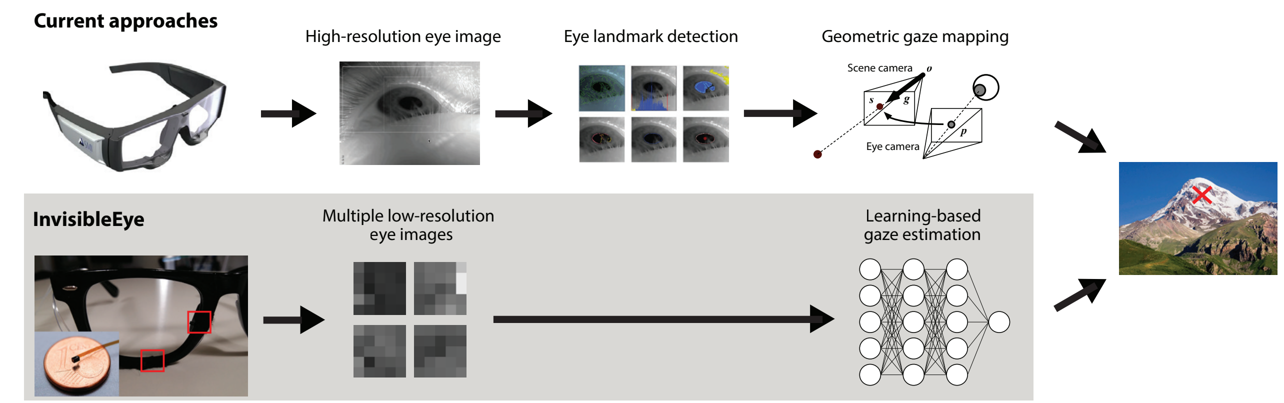


FIGURE 1. (Top) Classic approaches require high-resolution imaging sensors, resulting in rather bulky and obtrusive headsets, as well as hand-optimized algorithms for eye landmark detection and geometric gaze mapping. (Bottom) InvisibleEye is an innovative approach for mobile eye tracking that uses millimeter-size RGB cameras, which can be fully embedded into normal glasses frames. Our approach uses multiple cameras in parallel and appearance-based gaze estimation (red cross).

present three key experiments that we performed on different datasets and that show that our approach is competitive to state-of-the-art mobile eye trackers in terms of gaze estimation performance: The first dataset consists of 200,000 eye images synthesized using a recent computer graphics method [13] and allows us to explore the influence of the number of cameras, camera positioning, and image resolution on gaze estimation performance in a principled way. The second dataset contains 280,000 real eye images recorded with a first prototype implementation in a laboratory setting with controlled lighting. Finally, the third dataset has 240,000 real eye images recorded using a second prototype in a challenging unconstrained setting in which participants gazed at physical targets from various angles.

EXPERIMENT 1: Evaluation on Synthetic Images

The goal of the first experiment was to investigate the design space of fully embedded mobile eye tracking using synthetic eye image data, in particular, the minimum required number and positions of cameras.

Data Synthesis: The dataset for Experiment 1 was generated using *UnityEyes*, a computer graphics eye region model to synthesize highly realistic and perfectly annotated eye region images [12]. We synthesized images for five different eye regions as illustrated in Figure 2 (left). For each combination of eye region, camera angle, and lighting condition, we recorded a set of 1,600 different eyeball poses. Each set was randomly split into a set of 1,280 training images and 320 test images. To simulate the images that a low-quality sensor would yield, we down-sampled the images generated by UnityEyes to resolutions below 20x20 pixels. We converted them to grayscale to further lower their dimensionality.

Results: We trained different neural networks for different numbers of cameras. The results of this series of experiments are summarized in Figure 2 (middle). As can be seen from the figure, at a resolution of 5x5 pixels, our approach achieves a gaze estimation error of 0.084° using three and 0.073° using five cameras. Although the results achieved on synthetic data do not directly translate to the real world, given that the gaze estimation task is a lot easier without real-world noise, this first set of experiments clearly demonstrates that mobile gaze estimation does not necessarily require high-resolution images.

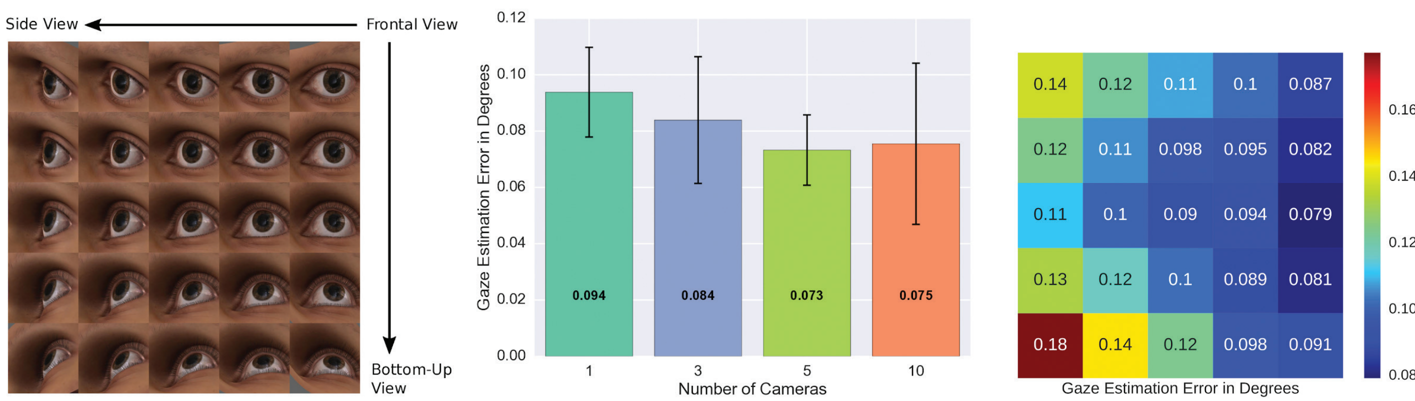
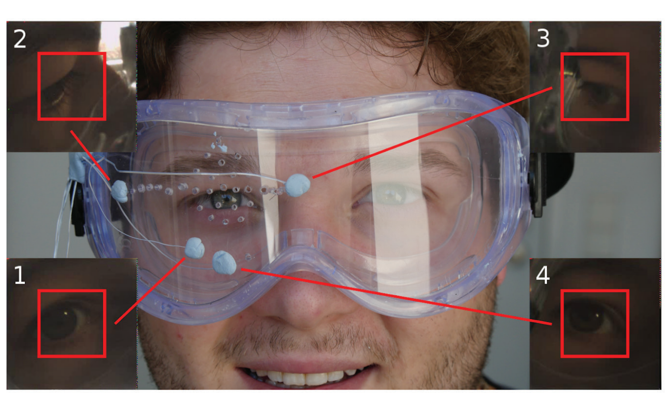


FIGURE 2. The use of synthetic images (left) allows us to evaluate the performance for a varying number of cameras at 5x5-pixel resolution (middle) as well as a wide range of camera angles (right) measuring the average gaze estimation in degrees.



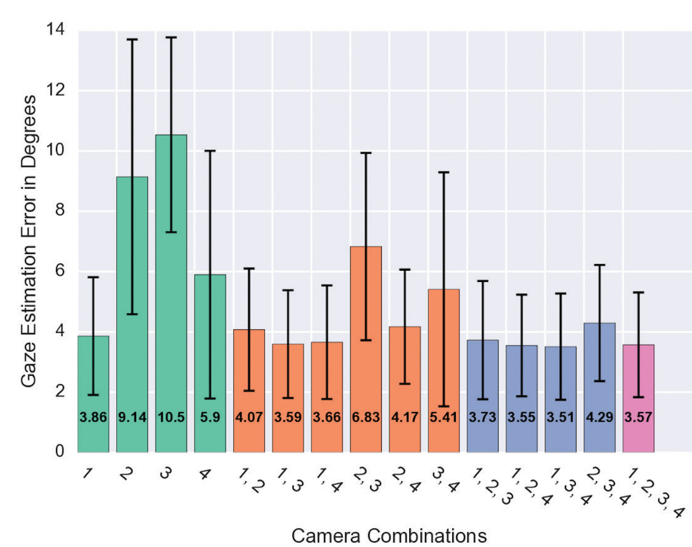


FIGURE 3. First prototype equipped with four NanEye cameras (left) to evaluate the average gaze estimation error for different camera combinations at 3x3-pixel resolution (right).

Another important parameter is the positioning of the cameras. Figure 2 (right) shows the error when using every available camera individually. As expected, frontal views of the eye yield the best results but are not viable in practice due to occlusion within the user's visual field of view. Since bottom-up views and pure side views were the next best options, we opted to position the cameras there in our prototype, which represents one of the key attributes and advantages of InvisibleEye.

EXPERIMENT 2: Evaluation in a Controlled Laboratory Setting

The goal of the second experiment was to evaluate a first hardware prototype of InvisibleEye on real images, but in a controlled laboratory environment. We opted for Awaiba NanEye cameras with a footprint of only 1x1 mm, an image resolution of 250x250 pixels, and 44 frames per second. The number of cameras and their positioning was informed by the first experiment.

Although the form factor of this medium-resolution camera is already sufficient to realize fully invisible mobile eye tracking (see Figure 1, bottom left), we also explored even lower image resolutions, i.e., below 20x20 pixels that promise further decreased bandwidth and computational requirements. We simulated this by reducing the image resolution manually.

The prototype was built by attaching four cameras to a pair of safety glasses (see Figure 3, left).

Data Collection: We used the prototype to record a second dataset of more than 280,000 close-up eye images with ground truth annotation of the gaze location of 17 participants (12 male, 5 female). For each participant, two sets of data were recorded: one set of training data and a separate set of test data. For each set, a series of gaze targets was shown on a display that participants were instructed to look at. A detailed description of the recording procedure can be found in [14] and the dataset at: *http://www.mpi-inf.mpg.de/invisibleeye*

Results: We computed the gaze estimation error of InvisibleEye for a resolution of 3x3 pixels varying camera combinations (see Figure 3, right). InvisibleEye is capable of estimating gaze with an error of 3.86° with a single camera. It achieves the lowest error of 3.51° with a combination of three cameras. This shows that gaze estimation at this low resolution is possible also with real-world data, which is sufficient for many practical applications like activity recognition [2] or attention analysis [11]. We also see that additional cameras do not help for every combination of cameras.

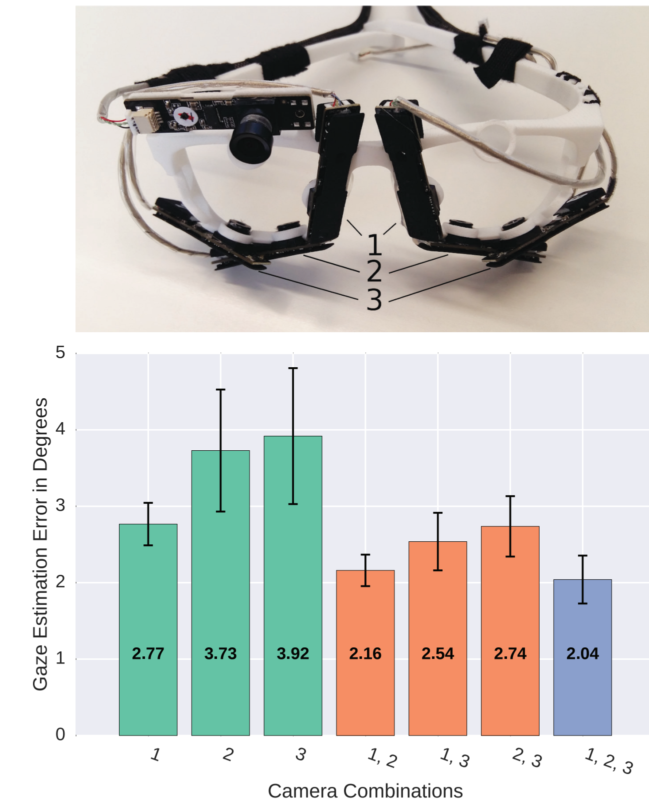


FIGURE 4. (Top) Second prototype consisting of custom 3D-printed glasses frame equipped with three camera pairs. (Bottom) Average gaze estimation error for different camera pair combinations at 3x3-pixel resolution.

INVISIBLEEYE IS AN INNOVATIVE APPROACH FOR MOBILE EYE TRACKING…ITS KEY FEATURE IS THE COMBINATION OF CAMERAS WITH A METHOD FOR APPEARANCE-BASED GAZE ESTIMATION

EXPERIMENT 3: Evaluation in an Unconstrained Setting

In the controlled setting, we assumed a display at a fixed distance in front of the user and predicted gaze in the screen coordinate system. For the final experiment, we built a second hardware prototype featuring a scene camera that records the user's field of view and allows us to test InvisibleEye in an unconstrained setting. We also explicitly allowed gaze targets at arbitrary depths. The depth at which a gaze target lies directly correlates with the location of the target projected into the camera image. From only the view of one eye, this location in the image is, however, in general not inferable. It is therefore necessary to use views from both eyes to resolve this ambiguity, which we do by using symmetric pairs of cameras recording both eyes. Further, we explicitly allow slippage of the headset, which is a problem frequently occurring in practice [10]. For this second prototype, we decided against using NanEye cameras to facilitate comparison with state-of-the-art mobile gaze estimation methods that require higher resolution images. We instead used Pupil Labs cameras [5] to record the eyes and the scene using a custom-built, 3D printed frame (see Figure 4, above).

Data Collection: Using this prototype, we recorded a third dataset of 240,000 eye images with four participants (four male, aged between 24 and 38 years). To record gaze data at varying distances, a calibration marker was attached to a wall in front of the participants. Participants were asked to position themselves at an arbitrary distance of up to 3 meters in front of the marker and to perform a series of head movements while gazing at the marker. The images recorded from each camera pair, i.e., one camera from the left side and its symmetrical counterpart from the right side, were concatenated.

Results: We first computed a baseline performance using a state-of-the-art gaze estimation approach based on pupil detection [5] on the original high-resolution images. This baseline method achieved an error of 10.96°. This high error is due to the strong slippage of the headset that is present in the data but not being compensated for. Similarly, as before, we evaluated the average gaze estimation performance of InvisibleEye for different camera pair combinations. In Figure 4 (previous page) we can see that, in all cases, the addition of a second camera pair improved the results on average for 3x3-pixel resolution. InvisibleEye achieves the best performance with an error of only 2.04° using all three camera pairs. These results demonstrate that InvisibleEye is a viable option even in the most challenging mobile settings.

CONCLUSION

InvisibleEye is an innovative approach that, in contrast to a long line of work on mobile eye tracking, relies on tiny cameras that can be nearly invisibly integrated into a normal glasses frame and, as such, addresses several key challenges of current systems. Its key feature is the combination of these cameras with a method for appearance-based gaze estimation. Results from our experiments not only underline the potential of this new approach but also mark an important step toward finally realizing the vision of fully unobtrusive, comfortable, and socially acceptable mobile eye tracking. ■

Julian Steil is a PhD student at the Max Planck Institute for Informatics in Germany. His current research focus encompasses the investigation of novel sensing approaches for mobile eye tracking devices, the analysis of human visual behavior in unconstrained settings, and the development of real-world applications for privacy-aware eye tracking and gaze-based activity recognition. jsteil@mpi-inf.mpg.de

Marc Tonsen is a research engineer at Pupil Labs GmbH. Since 2018, he has been co-leading the research and development team at Pupil Labs. He received his MSc in Computer Science from Saarland University. marc-tonsen@armun.de

Yusuke Sugano is an associate professor at the Institute of Industrial Science, the University of Tokyo. He received his PhD in Information Science and Technology from the University of Tokyo in 2010. His research interests focus on computer vision and human-computer interaction. sugano@iis.u-tokyo.ac.jp

Andreas Bulling is a professor at the University of Stuttgart, Germany, where he holds the chair for Human-Computer Interaction and Cognitive Systems. He received his MSc in Computer Science from the Karlsruhe Institute of Technology, Germany, and his PhD in Information Technology and Electrical Engineering from ETH Zurich, Switzerland. His research interests include human-computer interaction, computer vision, and machine learning. andreas.bulling@vis.uni-stuttgart.de

REFERENCES

[1] Andreas Bulling and Hans Gellersen. (2010). "Toward mobile eye-based human-computer interaction." *IEEE Pervasive Computing* 9, 4, 8–12. DOI: http://dx.doi.org/10.1109/MPRV.2010.86

[2] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Troster. (2011). "Eye movement analysis for activity recognition using electro-oculography." *IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 4* (2011), 741–753. DOI: http://dx.doi.org/10.1109/TPAMI.2010.86

[3] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. (2015). "Orbits: Enabling gaze interaction in smart watches using moving targets." In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*. 457–466. DOI: http://dx.doi.org/10.1145/2807442.2807499

[4] Sabrina Hoppe, Tobias Loetscher, Stephanie Morey, Andreas Bulling. (2018). "Eye movements during everyday behavior predict personality traits." Frontiers in Human Neuroscience, 12, 105:1-105:8. DOI: http://dx.doi.org/10.3389/fnhum.2018.00105

[5] Moritz Kassner, William Patera, and Andreas Bulling. 2014. "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction." In *Adj. Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 1151–1160. DOI: http://dx.doi.org/10.1145/2638728.2641695

[6] Päivi Majaranta and Andreas Bulling. (2014). *Eye Tracking and Eye-Based Human-Computer Interaction*. Springer Publishing London, 39–65. DOI: http://dx.doi.org/10.1007/978-1-4471-6392-3_3

[7] Evan F. Risko and Alan Kingstone. (2011). "Eyes wide shut: Implied social presence, eye tracking and attention." *Attention, Perception, & Psychophysics 73, 2* (2011), 291–296. DOI: http://dx.doi.org/10.3758/s13414-010-0042-1

[8] Hosnieh Sattar, Sabine Müller, Mario Fritz, and Andreas Bulling. 2015. "Prediction of search targets from fixations in open-world settings." In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 981–990. DOI: http://dx.doi.org/10.1109/CVPR.2015.7298700

[9] Julian Steil and Andreas Bulling. (2015). "Discovery of everyday human activities from long-term visual behavior using topic models." In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp). 75–85. DOI: http://dx.doi.org/10.1145/2750858.2807520

[10] Yusuke Sugano and Andreas Bulling. (2015). "Self-calibrating head-mounted eye trackers using egocentric visual saliency." In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*. 363–372. DOI: http://dx.doi.org/10.1145/2807442.2807445

[11] Yusuke Sugano, Xucong Zhang, and Andreas Bulling. (2016). "AggreGaze: collective estimation of audience attention on public displays." In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 821–831. DOI: 10.1145/2984511.2984536

[12] Erroll Wood, Tadas Baltrušaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling. (2016). "Learning an appearance-based gaze estimator from one million synthesized images. In *Proc. International Symposium on Eye Tracking Research and Applications (ETRA)*. ACM, 131–138. DOI: http://dx.doi.org/10.1145/2857491.2857492

[13] Erroll Wood, Tadas Baltrusaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling. (2015). "Rendering of eyes for eye-shape registration and gaze estimation." In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*. 3756–3764. DOI: http://dx.doi.org/10.1109/ICCV.2015.428

[14] Marc Tonsen, Julian Steil, Yusuke Sugano, and Andreas Bulling. (2017). "InvisibleEye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 1, 3, 106: 1–106:21. DOI: https://doi.org/10.1145/3130971

[15] Yanxia Zhang, Hans Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014. "GazeHorizon: Enabling passers-by to interact with public displays by Gaze. In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 559–563. DOI: http://dx.doi.org/10.1145/2632048.2636071