A Method to Compensate Head Movements for Mobile Eye Tracker Using Invisible Markers

Rie Osawa
The University of Tokyo,
Tokyo, Japan

Susumu Shirayama
The University of Tokyo,
Tokyo, Japan

Although mobile eye-trackers have wide measurement range of gaze, and high flexibility, it is difficult to judge what a subject is actually looking at based only on obtained coordinates, due to the influence of head movement. In this paper, a method to compensate for head movements while seeing the large screen with mobile eye-tracker is proposed, through the use of NIR-LED markers embedded on the screen. The head movements are compensated by performing template matching on the images of view camera to detect the actual eye position on the screen. As a result of the experiment, the detection rate of template matching was 98.6%, the average distance between the actual eye position and the corrected eye position was approximately 16 pixels for the projected image (1920 x 1080).

Introduction

Recently, various eye-tracking devices have been introduced to the market, and eye movement analyses is being conducted in many domains. The difference in gaze behavior between novices and experts can be utilized to develop efficient training methods (Vickers, 1996) (Klostermann et al., 2014). Also, the difference when changing color or arrangements of objects can also help for product development or marketing (Chandon et al., 2006; Wedel and Pieters, 2008).

There are generally two types of eye-gaze measurement devices, based on the pupil center corneal reflection method which uses near infrared (NIR) illuminators. One is a display installation type, where the NIRs are installed on a PC display to obtain the eye position. The other is a head mounted type, which obtains the coordinates through identifying the gaze position on a viewed image or movie.

In psychological studies, it is common for the subjects' heads to be fixed, in order to obtain accurate eye-movement measurements. However, in experiments to measure human gaze behavior realistically, restricting the subjects' head motion is far from the actual conditions, because humans are known to move their heads, consciously or unconsciously. Head motion represents one of the major human physiological behaviors and is essential in daily life (Hammal, 2014), which is why the decision against any motion restriction was made.

For eye tracking, a head-mounted type of device is suitable when considering reality and flexibility. However,

there is one problem specific of such devices: the output eye position data is affected by head movements. Sun et al. (2016) mention that it is important to remove noise such as head movements from the obtained gaze data in order to detect the degree of concentration of the driver. Therefore, several methods to detect the exact eye position while excluding the effect of head movements have been developed.

In our study, eye tracking is done utilizing a large screen with artificial feature points created by NIR-LEDs which cannot be seen by the naked human eye. Image processing is performed on the image of the view camera in which feature points are recorded, thereby compensating for the head movements. Finally, a method is proposed for automatic output of the exact part of the large screen being viewed by the subject.

Related Work

There are four major solutions that have been proposed to address the issue of matching eye positions in the view camera with actual eye positions on the screen or to compensate for head movements from entire eye-tracking data.

Methods based on features in images

Toyama et al. (2012) proposed using SIFT (Scale-Invariant Feature Transform) features. Points that have high contrast characteristics, or points at the corner, are regarded as key points with highly noticeable features. These are suitable for matching because they are not affected by rotation or scaling. Jensen et al. (2017) applied SIFT features to construct a 3D AOI (Area Of Interest) from eye-tracking data obtained by a head-mounted eye-tracker. Takemura et al. (2010) proposed using PTAM (Parallel Tracking and Mapping) and Chanijani et al. (2016) applied LLAH (Locally Likely Arrangement Hashing) to find feature points. However, these methods require a sufficient number of feature points in the image for accuracy, which may or may not be present depending on the contents of the view camera image.

Methods with markers

NAC Image Technology (2008) offers a method using AR markers to create artificial feature points, where AR markers captured in the view camera are matched with spatial coordinates. Tomi and Rambli (2016) also proposed using an eye-tracker with an AR application in the calibration of a head-mounted display. Huang and Tan

(2016) used circular patterns as markers. However, large markers could influence eye movements due to their size and appearance. Kocejko et al. (2014) proposed an algorithm to compensate for head movements with three cameras (to observe the eye, scene, and head angle) and LED markers. However, objects of view were limited in the monitor as were the movement of the subjects.

A method with infrared data communication

Tobii Technology offered a solution that uses infrared data communication markers. Eight such markers (approximately 30 mm^3) are required for position detection, where each marker communicates with the eye-tracking device and matches the image of the view camera with the respective spatial coordinates. However, the size of such markers could have a significant impact on eye movements. Note that this device is not currently available.

Methods sensing head movements

Ahlstrom et al. (2012) proposed compensating for head movements using the recorded gaze behaviors in actual driving scenes with a video camera. However, detection of head movement is performed manually for each frame, which is costly. Larsson et al. (2014) applied a gyro, an accelerometer, and a magnetometer. Even though the accuracy has been improved, the synchronization of eye-tracking data and other sensors still remains an issue.

Proposed Methodology

As mentioned above, methods based on features require a sufficient number of feature points in the image for accuracy. Markers could influence eye movements in the method with AR markers or infrared data communication. In the method sensing head movements, the synchronization of eye-tracking data and other sensors still remains an issue. In our proposed methodology, eye tracking is done utilizing a large screen with artificial feature points created by NIR-LEDs which cannot be seen by the naked human eye. This methodology does not rely on the content of visual stimuli therefore can be applied even when there are not sufficient features there. Furthermore, markers created by NIR-LEDs does not affect eye movements. Image processing is performed on the image of the view camera in which feature points are recorded, thereby compensating for the head movements. Since template matching is automatically performed using image processing, cost is low

and processing is fast relatively. Finally, a method is proposed for automatic output of the exact part of the large screen being viewed by the subject.

Overview of the experimental apparatus

Figure 1 illustrates the overview of the experimental apparatus devised to measure the subjects' gaze behavior while watching the large screen. Figure 2 illustrates the actual experimental environment.
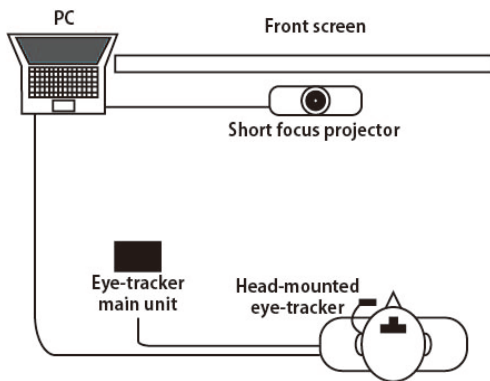

Figure 1: Overview of the experimental apparatus.

Figure 2: Experimental apparatus for eye tracking.

Eye-tracking apparatus

To record the data of the eye position, we selected NAC Image Technology's EMR-9 as the eye-tracking device, which includes a view camera attached to the subject's forehead for video recording. The eye position is indicated by the x–y coordinates in the area recorded by the view camera (Figure 3). Even if the eye position is fixed on a specific item, head movements will cause shifts in the

view camera area and the x–y coordinates, leading to difficulties in identifying the target object, as seen in Figure 4.

Figure 3: Output image of the view camera.

New method using artificial feature points with infrared LED markers

In this paper, a new eye-tracking method is proposed via the creation of artificial feature points made of invisible NIR (near-infrared)-LED markers and image processing. NIR-LEDs are invisible to the human naked eye, therefore reducing their effect on eye tracking despite their presence. At the same time, NIR-LEDs are visible through IR filters, as seen in Figure 5. In robot technology, it is popular to use NIR-LEDs to detect locations or to follow target objects (Sohn et al., 2007). However, to the authors' best knowledge, there have been no NIR-LED applications used for eye tracking, which has the potential to enable eye-movement detection even with head movements.

The view camera with IR filters captures the feature points of NIR-LEDs installed on the projection screen. This image can be used to verify the eye position relative to the NIR-LED feature points, which can then be used to calculate exactly what the subject is looking at on the screen by image processing. We call these invisible NIR-LED markers "IR markers" hereafter.

Image processing is another question that requires attention. SIFT features could be a potential option. However, these methods are not adequate for images of IR markers received through the IR filter, because single NIR-LED IR markers are homogeneous and less characteristic, as shown in the image on the right side of Figure 5. As a countermeasure, several patterns composed of mul-

tiple NIR-LEDs have been developed as matching templates, as described below. The overall flow is described later.

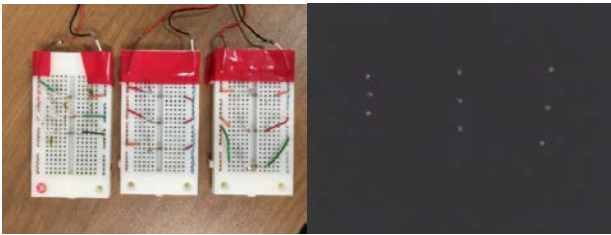Figure 4: Variations in the view and axis caused by head movements.

Figure 5: IR markers with the naked eye (left) and through the filter (right).

Patterns of IR markers

IR Marker patterns have been created taking into account the four following conditions.

1. Patterns should have a sufficient number of features.

2. Patterns should be composed of the smallest number of markers possible.

3. Patterns should be sufficiently differentiable from one another.

4. Patterns should be easily produced.

To decide on the exact patterns, the similarities between patterns of filtered IR markers (Figure 6) have been schematically calculated. Taking condition 2 into consideration, a three-point pattern was selected from a 5×5 dot matrix for each pattern, which was the best balance to ensure noticeable differentiation. Similarities are calculated by Hu invariant moment algorithm (Hu, 1962).

For a two-dimensional continuous function $f(x, y)$, the moment of order $(p + q)$ is defined as Equation (1).

$$m_{pq} = \sum_{x,y} x^p y^q f(x, y) \tag{1}$$

The image moment is the variance value of the pixel centered on the origin of the image. Here, the suffix represents the weight in the axial direction. Subsequently, the centroid is obtained by Equation (2).

$$(\bar{x}, \bar{y}) = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{2}$$

The pixel point $(\bar{x}, \bar{y})$ are the centroid of the image $f(x, y)$. Based on the coordinates of this centroid, the moment considering the centroid is obtained by Equation (3).

$$\mu_{p,q} = \sum_{x,y} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \tag{3}$$

Further, normalize this moment of centroid by Equation (4) to find the normalized centroid.

$$\eta_{ij} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{4}$$

where

$$\gamma = \frac{i + j}{2} + 1, i + j \geq 2$$

By normalizing, the variance no longer affects the moment value, therefore it is invariant to the scale.

Seven kinds of Hu invariant moment are defined by using the normalized centroid moment, in this study, the moment is calculated by Equation (5).

$$\varphi = \eta_{20} + \eta_{02} \tag{5}$$

This is the sum of variances in the x-axis direction and the y-axis direction.

Table 1 shows the result of the similarity calculation using the Hu invariant moment.

The template images are on the top of the table and the searched images are on the side of the table; lower matching evaluation scores indicate higher similarity and are represented with red cells. The Hu invariant moment allows checks of both rotational and scale invariance; therefore, relevant combinations of patterns with high similarity scores can be calculated.
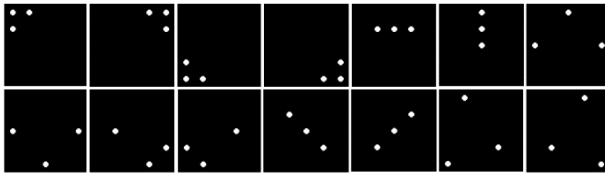

Figure 6: Patterns used for the experiment.

Table 1: The result of the matching experiment.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		0.002	0.005	0.039	0.077	0.079	0.415	0.416	0.308	0.301	0.203	0.205	0.452	0.453
2			0.003	0.041	0.035	0.038	0.307	0.308	0.212	0.203	0.119	0.120	0.367	0.367
3				0.043	0.037	0.040	0.310	0.311	0.214	0.206	0.121	0.122	0.370	0.370
4					0.006	0.003	0.267	0.268	0.171	0.162	0.078	0.079	0.326	0.327
5						0.004	0.338	0.340	0.231	0.225	0.126	0.128	0.375	0.377
6							0.336	0.337	0.229	0.222	0.124	0.126	0.373	0.374
7								0.002	0.148	0.150	0.212	0.210	0.107	0.105
8									0.150	0.152	0.213	0.212	0.106	0.105
9										0.016	0.105	0.103	0.232	0.231
10											0.098	0.097	0.238	0.237
11												0.002	0.249	0.250
12													0.247	0.249
13														0.001
14														

Based on the findings, several patterns were chosen and created with IR markers. Specifically, NIR-LEDs and resistors were attached to a solder-less breadboard and were mounted onto a polystyrene board. To ensure the high accuracy of the template matching, it was found that twelve patterns were required to be on the board for at least three patterns to be within the view camera at a given time for image processing. The layout of the IR markers was decided based on the similarity results seen in Table 1, and the actual implementation can be seen in Figure 7.
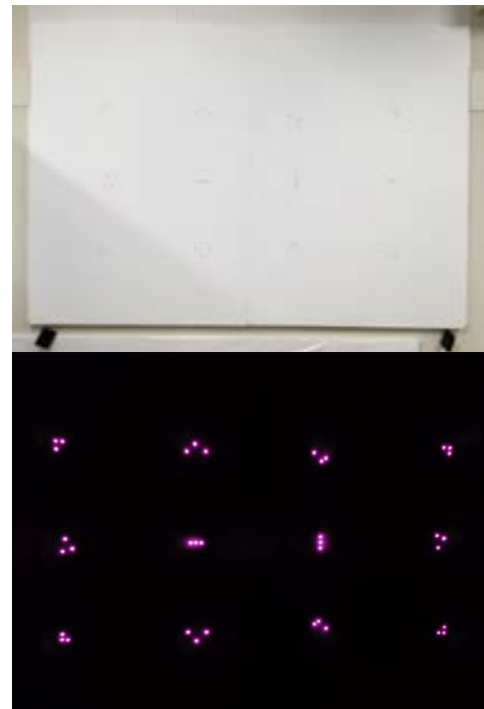

Figure 7: IR Marker-embedded screen as seen with the naked eye (top) and through a filter (bottom).

Procedure

The operation principle and pattern creating method of NIR-LEDs are described in the previous section. In this section, we will introduce the process and the algorithm of calculating the subject's view point on the screen, derived from the LED points on the screen and the eye positions.

1. Distortion of the image is caused by the lens of the view camera, therefore calibration is performed for each frame of the obtained movie.

2. Apply template matching on the distortion-corrected images of the view camera to detect the IDs of the IR markers and their coordinates.

3. Detect three points with high matching rates, and obtain their coordinates. In order to calculate the line-of-sight positions on the screen, apply affine transformation to the known coordinates of the markers on the screen.

4. Map the corrected eye coordinates on the image projected on the screen (Figure 8).

5. Output the image or movie with the mapped eye positions (format depends on the visual source).

Affine transformation is used to map coordinates of eye positions in the images of view camera onto the screen. Specifically, scaling is required to adjust the difference in resolution between the view camera and the image projected on the screen, rotation and translation are required to compensate the head movements. Affine transformation is a movement and deformation of a shape that preserves collinearity, including geometric contraction, expansion, dilation, reflection, rotation, shear, similarity transformations, spiral similarities, translation and compositions of them in any combination and sequence.

These transformations for point $p(x, y)$ on a plane to be mapped to point $p'\left(x', y'\right)$ on another plane are expressed as Equation (6).

$$p' = f(p) = Ap + t \tag{6}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad t = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

A represents a linear transformation, and *t* represents a translation. Scaling can be expressed as Equation (7).

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{7}$$

α and β are scale factors of *x*-axis and *y*-axis direction respectively. Similarly, rotation can be expressed as Equation (8).

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{8}$$

θ is the angle of rotation in the mapped plane. Scaling, rotation and translation are used in this research because distortion caused by the lens of the view camera is calibrated before affine transformation, and scale factor is common to *x*-axis and *y*-axis. Therefore, affine transformation matrix required to detect eye positions are obtained by Equation (9).

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b_1 & b_2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{9}$$

Figure 8 illustrates the image of affine transformation used in our method.
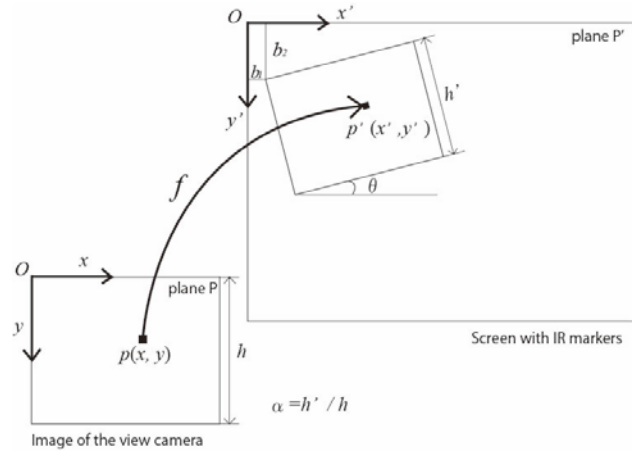

Figure 8: Affine transformation between the image of the view camera and the screen.

Verification experiment

Implementation of the screen for eye tracking

Verification experiments were conducted to examine the proposed method's correlation between the eye position, as seen through the view camera, and the actual projected image. Since our method assumes covering the field camera with a filter, the image from the view camera won't allow detection of what the subject is looking at. In order to verify the results, template matching was conducted by creating a simulated filtered image, by projecting an identical image of that seen on the view camera onto the screen through an IR filter. The image projected on the screen is shown in Figure 9.

Preliminary experiment

Before conducting template matching of all gaze data, preliminary experiments were conducted to confirm template matching performance. The numbers 1 through 3 were added to the image seen in Figure 9 and projected as shown in Figure 10, where the subjects wearing the EMR-9 eye tracker were requested to look at them in order. Figure 11 shows an image clipped from the view camera movie during eye-tracking measurements, and Figure 12 represents six template matching results with obtained gaze data.
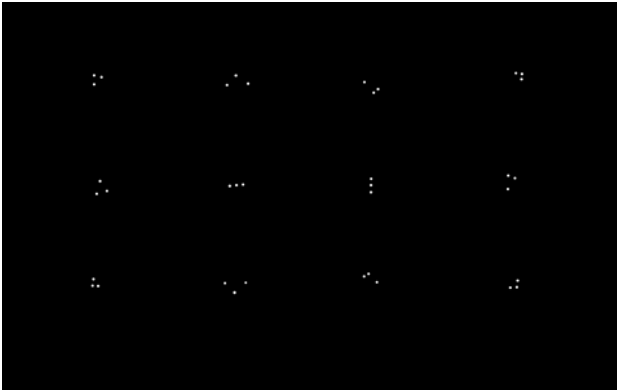
Figure 9: Projected image on the screen for verification experiment.
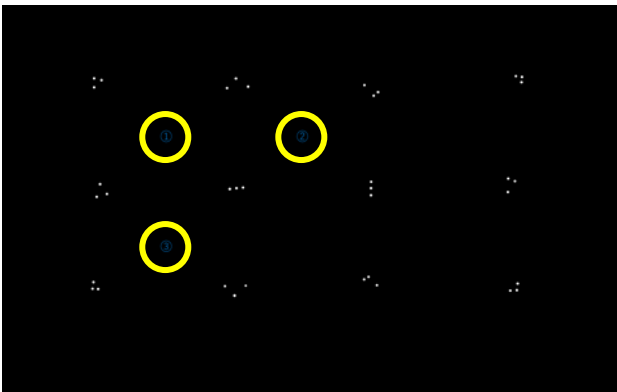

Figure 10: Image source projected on the screen for preliminary experiment (top; yellow circles added for enhancement) and screen with image source projected (bottom).

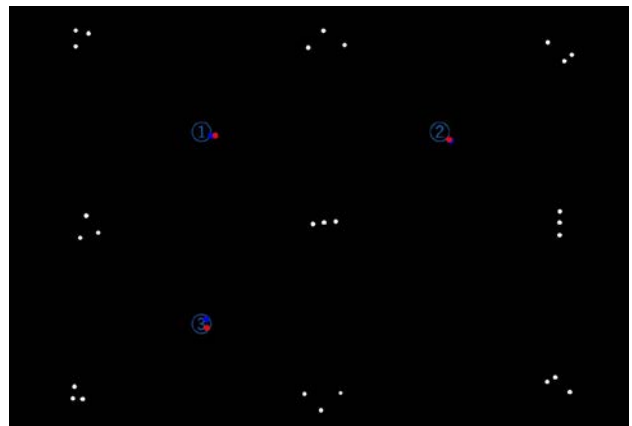Figure 11: Captured image of the view camera while the subject watching the number 1 on the screen.

Figure 12: Result of preliminary experiment (enlarged). Eye positions while watching the number 1 through 3 plotted on the projected image through the template matching.

Result of gaze plot

Gaze behaviors of the subjects were measured with EMR-9 at 30fps, in a zigzag manner from the upper left marker to the lower right marker of the image shown in Figure 9. Subjects could move their heads freely. To verify template matching performance, Affine transformation was manually conducted based on the template shown in the view camera's image, and eye positions were mapped onto the projected image. Figure 13 represents template matching results, including a comparison with manually mapped eye points.

Approximately 250 eye points were mapped, where data suggests a very high correlation between template

matching and manually conducted mapping results, although some deviation does remain.

Let Δd_i be the distance between the actual eye position and the corrected eye position, where i denotes the ith eye points. The detection rate of template matching was 98.6%. Averaged Δd was 15.9 pixels. Note that the resolution of the projected image was 1920 x 1080 pixels. Points containing detection errors can be seen in Figure 14 and the histogram of Δd_i is represented in Figure 15. More than 90% of Δd_i are within 30 pixels. The main cause of such errors is due to view camera image capture failures, caused by very quick head motions and camera shake, leading to image blur which prevents accurate template matching. However, for example, Δd_i of 30 pixels falls within the range of rear combination lamp of the car shown in the top of Figure 17 (a white circle at point A represents 30 pixels). It can therefore be assumed that our method works in practical use.

Gaze plot on the movie

Our method can also be used for gaze measurement while watching a movie, and output the movie with eye point mapped on each frame automatically. Here, a driving video footage taken from the inside of a vehicle while driving was adopted as a visual stimulus. Figure 16 shows the images clipped from the movie and Figure 17 shows the images of view camera and the corresponding corrected eye positions mapped on the source movie.

Conclusions and Remarks

A new method to compensate for head movement during eye-tracking has been developed, using invisible markers. This will enable higher eye position detection accuracy, which is a problem specific to mobile eye-tracker. However, our methodology has two limitations: First is that the eye tracking is limited on the screen with IR markers embedded. When expanding the range of measurement, it is necessary to add new screens and increase the number of markers newly. Secondly, current apparatus does not allow to confirm the correspondence between the projected image and the eye position in the image of view camera because the view camera is covered with the IR filter. In order to solve this issue, we will add a view camera without filter in the future work.

In addition to the issues to be solved in the future works shown above, error in positioning still remains, due to the error of template matching in some cases, which does have room for improvement for better eye position recognition. Potential solutions to reduce such error include (i) the use of a view camera with higher sensitivity and resolution with shorter exposure time, and (ii) adopting a more robust template matching method. As (i) is less realistic due to the wide use of commercially available eye-trackers with limited performance, a more effective approach would be (ii) through image preprocessing with edge detection as an example.
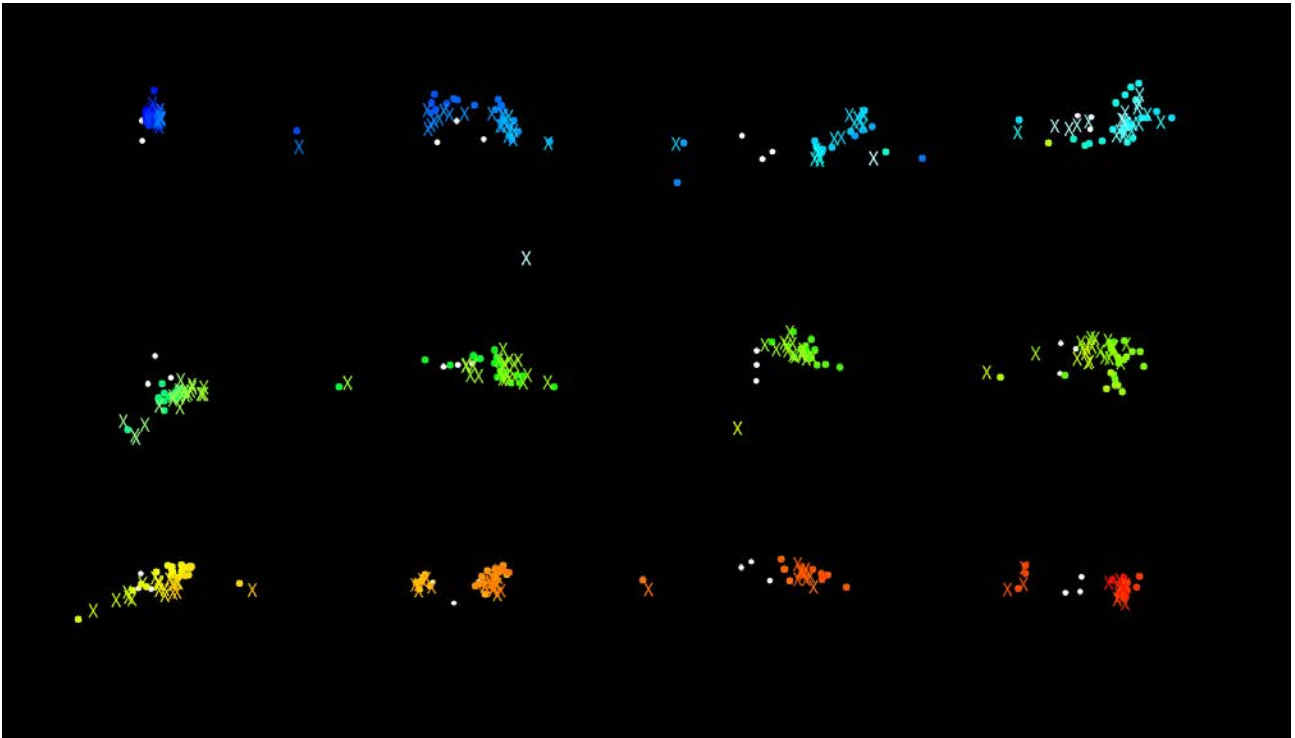
Figure 13: Projected image with eye position mapped (circles: results of template matching, X: results of manual mapping)
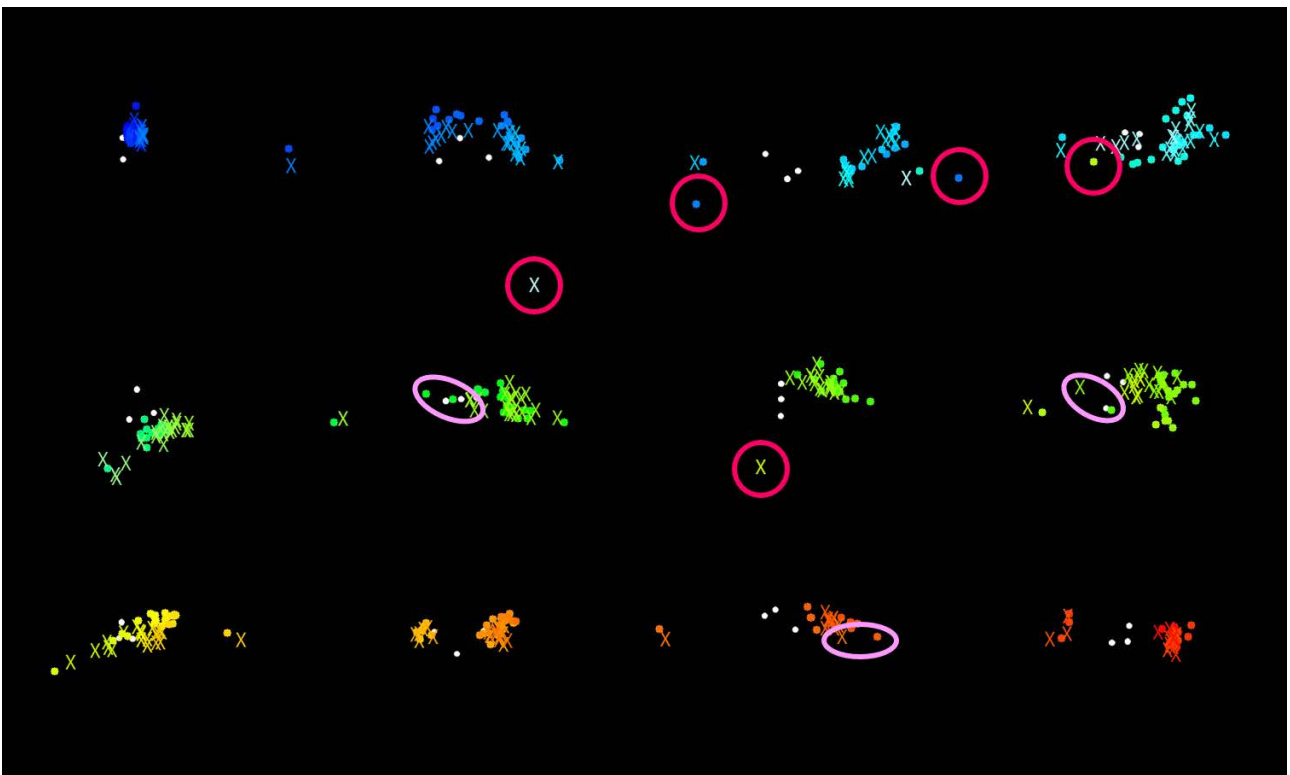

Figure 14: Points containing the errors of template matching (light pink: relatively large gap, dark pink: no correspondence)
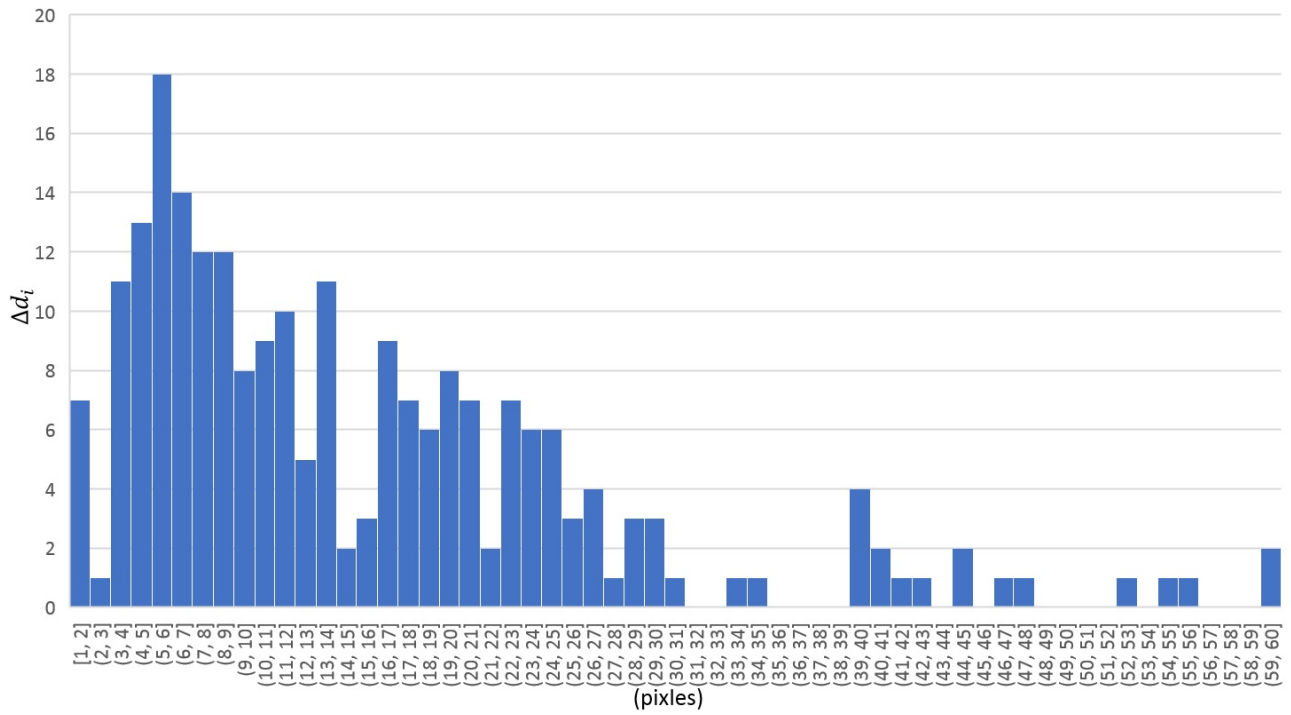
Figure 15: Histogram of distance between the actual eye position and the corrected eye position (Δd_i).

Figure 16: Scene images from a projected movie on the screen.

Figure 17: Examples of the images of filtered view camera (left; pink dots: eye positions) and the corresponding corrected eye position mapped on the video footage (right; corrected eye positions are circled).

References

Ahlstrom, C., Victor, T., Wege, C., & Steinmetz, E. (2012). Processing of eye/head-tracking data in large-scale naturalistic driving data sets, *IEEE Transactions on intelligent transportation system*, *13*(2), 553-564. Retrieved from http://dx.doi.org/10.1109/tits.2011.2174786

Chandon, P., Hutchinson, J. W., Bradlow, E., & Young, S. H. (2006). Measuring the value of point-of-purchase marketing with commercial eye-tracking data. *INSEAD Business School Research Paper, 2007/22*/MKT/AC-GRD. Retrieved from http://dx.doi.org/10.2139/ssrn.1032162

Chanijani, S. S. M., Al-Naser, M., Bukhari, S. S., Borth, D., Allen, S. E. M., & Dengel, A. (2016). An eye movement study on scientific papers using wearable eye tracking technology, *9th International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*. Retrieved from http://dx.doi.org/10.1109/icmu.2016.7742085

Hammal, Z., & Cohn, J. F. (2014). Intra- and interpersonal functions of head motion in emotion communication, *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges (RFMIR)*, 19-22. Retrieved from http://dx.doi.org/10.1145/2666253.2666258

Huang, C. W., & Tan, W. C. (2016). An approach of head movement compensation when using a head mounted eye tracker, *International Conference of Consumer Electronics-Taiwan (ICCE-TW)*. Retrieved from http://dx.doi.org/10.1109/icce-tw.2016.7520987

Hu, M. K. (1962). Visual pattern recognition by moment invariants, *IRE Transactions on Information Theory*, *8*(2), 179-187. Retrieved from http://dx.doi.org/10.1109/tit.1962.1057692

Jensen, R.R., Stets, J.D., Suurmets, S., Clement, J., & Aanas, H. (2017). Wearable gaze trackers: mapping visual attention in 3D, *Lecture Notes in Computer Science, 10269*, 66-76. Retrieved from http://dx.doi.org/10.1007/978-3-319-59126-1_6

Klostermann, A., Kredel, R., & Hossner, E-J. (2014). On the interaction of attentional focus and gaze: the quiet eye inhibits focus-related performance decrements, *Journal Sport and Exercise Psychology*, *36*, 392-400. Retrieved from http://dx.doi.org/10.1123/jsep.2013-0273

Kocejko, T., Bujnowski, A., Ruminski, J., Bylinska, E., & Wtorek, J. (2014). Head movement compensation algorithm in multi-display communication by gaze, *7th International Conference on Human System Interactions (HSI)*, 88-94. Retrieved from http://dx.doi.org/10.1109/hsi.2014.6860454

Larsson, L., Shwaller, A., Holmqvist, K., Nystrom, M. & Stridh, M. (2014). Compensation of head movements in mobile eye-tracking data using an inertial measurement unit, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1161-1167. Retrieved from http://dx.doi.org/10.1145/2638728.2641693

NAC Image Technology. (2008) EMR-dStream: Retrieved from http://www.eyemark.jp/product/emr_dstream/

Sun, Q., Xia, J., Falkmer, T., & Lee, H. (2016). Investigating the spatial pattern of older drivers' eye fixation bahaviour and associations with their visual capacity. *Journal of Eye Movement Research, 9*(6):2, 1-16. Retrieved from http://dx.doi.org/10.16910/jemr.9.6.2

Sohn, B, Lee, J., Chae, H. & Yu, W. (2007). Localization system for mobile robot using wire-less communication with IR landmark, *Proceedings of the 1st International Conference on Robot Communication and Coordination*, 1-6. Retrieved from http://dx.doi.org/10.4108/icst.robocomm2007.2173

Takemura, K., Kohashi, Y., Suenaga, T., Takamatsu, J., & Ogasawara, T. (2010). Estimating 3D point-of-regard and visualizing gaze trajectories under natural head movements, *Proceedings of 6th ACM Symposium on Eye Tracking Research & Applications (ETRA)*, 157-160. Retrieved from http://dx.doi.org/10.1145/1743666.1743705

Tomi, A. B., & Rambli, D. R. A. (2016). Automated calibration for optical see-through head mounted display using display screen space based eye tracking, *3rd International Conference on Computer and Information Science (ICCOINS)*, 448-453. Retrieved from http://dx.doi.org/10.1109/iccoins.2016.7783257

Toyama, T., Kieninger, T., Shafait, F., & Dengel, A. (2012). Gaze guided object recognition using a head-mounted eye tracker, *Proceedings of 7th ACM Symposium on Eye Tracking Research & Applications (ETRA)*, 91-98. Retrieved from http://dx.doi.org/10.1145/2168556.2168570

Wedel, M., & Pieters, R. (2008). Eye tracking for visual marketing, *Foundations and Trends in Marketing, 1*(4), 231-320. Retrieved from http://dx.doi.org/10.1561/1700000011

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of Experimental Psychology Human Perception and Performance*, *22*(2), 342-354. Retrieved from http://dx.doi.org/10.1037/0096-1523.22.2.342