Improving Eye Tracking Accuracy by Head Motion History

Masahiro TOYOURA† (*Member*), Takumi TANAKA†, Atsushi SUGIURA†, Xiaoyang MAO† (*Member*)

† University of Yamanashi

<**Summary**> We focus on eye tracking by head motion. This type of eye tracking does not provide the most accurate results, but it does not require a user to wear cumbersome sensors like cameras on glasses. The approach works for many applications, such as the extraction of human attention by surveillance camera or an intuitive interface for tablet devices. Through a preliminary experiment, we confirmed that head direction is often largely different from eye direction. We propose to estimate accurate eye direction by using head motion history. A sequence of head directions and the differentials provide richer information than head direction at one moment. Using multiple regression analysis (MRA) and dynamic coupled component analysis (DCCA), we examined the relationship between eye direction and head motion history, and reduced the error rate by 7.2% and 0.8% on average.

Keywords: eye tracking, ego-motion, gyro sensing, visual saliency

1. Introduction

Eye direction is a promising cue for understanding a user's attention. For example, eye tracking data obtained by surveillance camera images detect human attention for public signs. The most common technique for estimating eye direction is capturing images of human eyes[1]. Cameras installed on a person's glasses or at the frontal side of a monitor display are user for the purpose. However, this method forces a user to wear cumbersome devices, in case of cameras on glasses, or to sit on a chair, in case of cameras on a monitor display. To understand human attention in a real context, the method is not applicable. In contrast, eye tracking by head motion can be applied for crowds on a street. It can be used on infant subjects who hesitate to wear glasses or other special devices. However, eye movement estimated with head motion is not as accurate as tracking eye movement with cameras. Sankaranarayanan et. al. proposed an eye tracking method for eye tracking of pedestrians captured in multiple surveillance camera video images[2]. They assumed that a person's face direction is the same as his or her eye direction. The assumption is often used for such applications, but they have not focused on its accuracy.

In this paper, we focus on improving the accuracy of eye tracking data by head motion. This is a pilot study of using head motion history, not only head direction at a certain moment. We employ a gyro sensor on the head for data logging. **Figure 1** shows experimental environment. The gyro sensor will be replaced by a surveillance camera or other lighter devices in future works; however, this is not the focus of this paper. Also note that head motion history cannot represent eye direction. It is clear that humans can change eye direction without moving their heads. The contribution of our method is to improve the accuracy of eye tracking by head motion history.

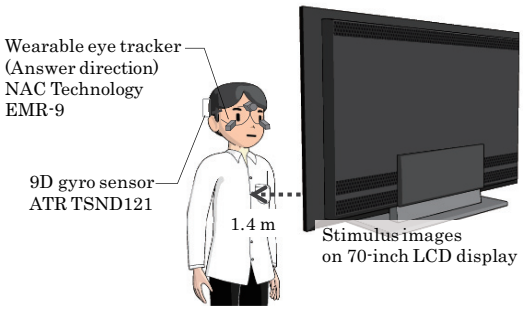


Fig. 1 Environment of experiment

When the directions of the head and the chest are the same, eye direction estimated only from head direction at the moment is the same as the direction of the chest. If the head is directed to the right at the previous moment, we can expect that the eye will direct to the left at the next moment. Humans see an object by collaboratively moving their eyes and heads. It is called "eye-head coordination"[3]. This topic is well discussed in the context of cognitive psychology. Many experiments have been conducted within limited components such as displacement from the center position. The difference of

individuals should be taken into account for engineering applications, but it tends to be ignored. In the context of computer vision or motion estimation, Okinaka et al.[4] also focused on eye-head coordination, and tried to improve the accuracy of eye tracking. They employed the speed and acceleration of head direction. Our employed head motion history includes the trajectory of the head direction in addition to speed and acceleration. A more sophisticated regression model improves the accuracy.

We have proposed a method for improving the accuracy of eye tracking by head motion in[5]. We further discuss on which stimulus image contributes to more improve the accuracy of eye tracking in this paper. The distribution of salient regions in the stimulus image tends to affect the degree of improvement.

We first examine the accuracy of eye tracking by head motion in Section 2. The method for improving the accuracy of eye tracking by head motion history is represented in Section 3. Experimental results are shown in Section 4, and we conclude the paper and discuss future work in Section 5.

2. Accuracy of Eye Tracking

We first examined the accuracy of eye tracking. As shown in Figure 1, still stimulus images were displayed on a 70-inch LCD display monitor. A subject stood 1.4m away from the monitor. Our employed eye tracker NAC Technology EMR-9 estimates eye direction with an accuracy of 0.1 degrees. The eye tracker has three cameras capture images of two eyeballs and outward. When assuming the directions of eyes and head are the same, the eye points always come to the center of images captured by the camera that shoots outward. The distance between the eye position and the center of an image is equal to the error of eye tracking by head direction at a given moment.

An example of root mean square error (RSME) of the angular error is plotted in **Figure 2**. The average is about 30 degrees, and the maximum is more than 70 degrees. The accuracy is much worse than with eye trackers on glasses.

3. Eye Tracking by Head Motion History

We tried to improve the accuracy of eye tracking. The core idea was using head motion history, not only head direction at a moment. As shown in **Figure 3**, the direction of head is not same as the direction of eyes especially in motion. The human eye-head coordination is known

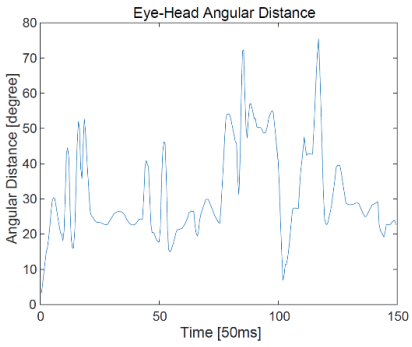


Fig. 2 Error of eye tracking by head motion

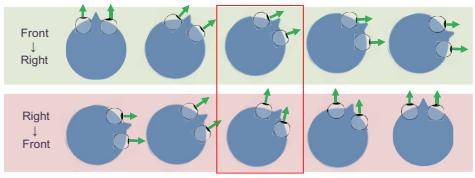


Fig. 3 Difference of eye direction from different head motion history

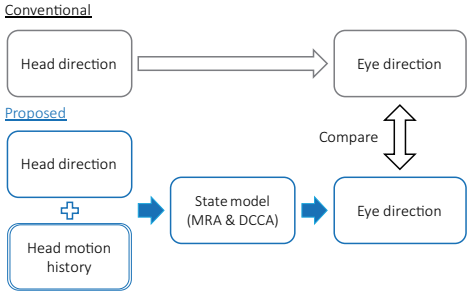


Fig. 4 Improving the accuracy of eye tracking by head motion history

as compensating movement and synergistic movement[6]. Related to the moment of eye movement, the compensating movement of head is observed between -240ms to +80ms, and the synergistic movement is observed -240ms to 0ms. The head motion history provides a cue for correctly estimating the eye direction.

Figure 4 shows the overview of this paper. The conventional model treated the eye direction as the head direction. Our proposed method introduces head motion history and two state models for integrating eye and head direction. The head direction and head motion history are logged by wearable devices in the experiment. Although we understand the wearable devices are not realistic for a practical use, since our focus is to clarify the possibility of head motion history for accurate eye tracking, we keep the problem as future work. We compare the accuracy of eye directions estimated with the conventional method and proposed method.

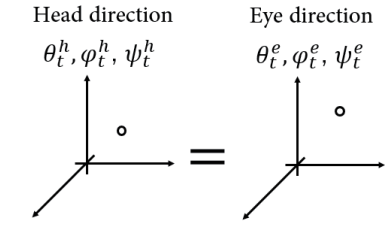
We used multiple regression analysis (MRA) and dynamic coupled component analysis (DCCA)[7]to learn the

relationship between head motion history and eye direction. MRA has a simple model, so it is easy to use, but it is difficult to learn non-linear angular relationships with MRA. DCCA elucidates the relationships between different dimensional datasets. It does not directly project the input vector into the output vector. DCCA assumes a lower dimensional space substantively controls the input and output, and shows the projections of input and output vectors for the lower dimensional space. DCCA works well even though the input and output vectors are represented in a high dimensional spaces. DCCA is robust for non-linear data, compared with simple MRA. Ma and Deng[8]applied DCCA to synthesize avatar models with eye motion. They employed DCCA to control eye motion of the avatar by human head tracking data. The introduction of smoothness function makes it robust even for noisy data. A disadvantage is the lower speed of convergence because of the complexity of the model.

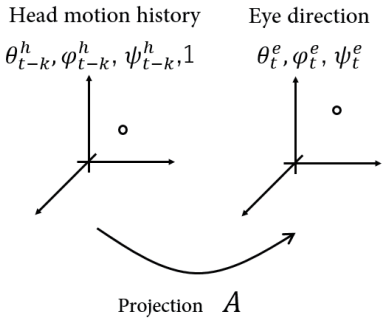
Figure 5 shows the difference among three types of learning. The input vector is roll θ^h_{t-k}, pitch ϕ^h_{t-k} and yaw ψ^h_{t-k} of head direction at time t, its previous time $t-k$ $(k=0,\cdots,n)$ and constant 1. Total dimension is $3(n+1)+1$. The output vector is roll θ^e_t, pitch ϕ^e_t, and yaw ψ^e_t of eye direction at time t. The dimension is 3. In the conventional model shown in Figure 5(a), eye direction is estimated as the same direction of the head. MRA, in Figure 5(b), shows the relationship between the input and output vectors as a matrix that projects the input vector into the corresponding output vector. DCCA, in Figure 5(c), shows the relationship by two matrices that project the input into the corresponding lower dimensional point, and the point is projected into the output with small error. In the learning phase, we gathered the input and output vectors with the glasses-type eye tracker and the gyro sensor mounted on the head. By MRA, a projection matrix A is optimized to minimize energy E defined by the input vectors and output vectors as the following.

$$e_t = \begin{pmatrix} \theta^e_t \\ \phi^e_t \\ \psi^e_t \end{pmatrix}, h_t = \begin{pmatrix} \theta^e_{t-n} \\ \phi^e_{t-n} \\ \psi^e_{t-n} \\ \vdots \\ \theta^e_t \\ \phi^e_t \\ \psi^e_t \\ 1 \end{pmatrix},$$

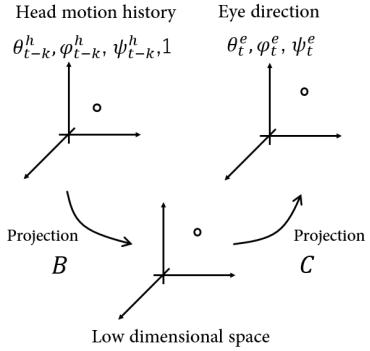
$$E = \sum^{t} ||e_t - Ah_t||. \quad (1)$$



(a) Conventional model



(b) Multiple regression analysis (MRA)



(c) Dynamic coupled component analysis (DCCA)

Fig. 5 Differences among three learning models

DCCA treats the projection matrices B and C as follows. S is a temporal smoothing matrix. The matrices B and C contribute to project the signal to lower dimensional space, and the matrix S contributes smoothing of the signal. B and C are essential, and S is complemental in DCCA. The hat notations mean that the dimensions are reduced by principle component analysis (PCA) or other techniques. λ_1 and λ_2 are arbitrary constants for weighting the terms.

$$E_{static} = \sum^{t} \left(||\hat{h}_t - Ce_t||_2^2 + \lambda_1 ||e_t - Bh_t||_2^2 \right), \quad (2)$$

$$E_{dynamic} = E_{static} + \lambda_2 \sum^{2} ||e_t - Se_{t-1}||_2^2. \quad (3)$$

E_{static} represents the energy that the input vectors and output vectors are projected, to the same corresponding points of the common lower space, and $E_{dynamic}$ controls the temporal smoothness of the output vectors.

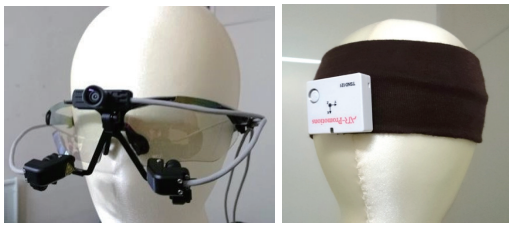
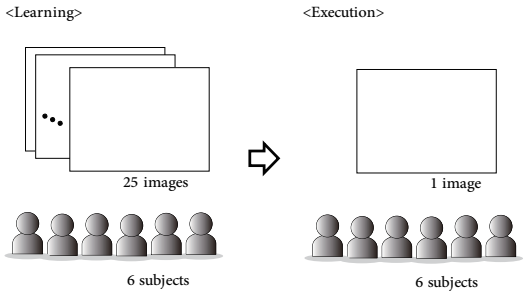
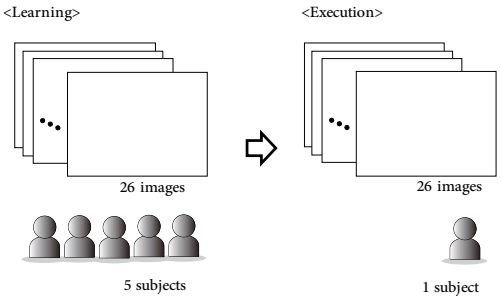


Fig. 6 Wearable eye tracker and 9D gyro sensor for the experiment. Subjects stare at the stimulus images with wearing the devices.



(a) Inter-user study



(b) Intra-user study

Fig. 7 Design of subject studies

4. Experiments

We conducted a subject study to investigate whether the proposed method improves accuracy. Six subjects (five males and one female; university students in their 20s) stood 1.4m away from a 70-inch monitor, as shown in Figure 1. A 9D gyro sensor was mounted on their heads, and they wore an eye tracker as shown in **Figure 6**. The equipment recorded head motion history and eye directions in time sequence.

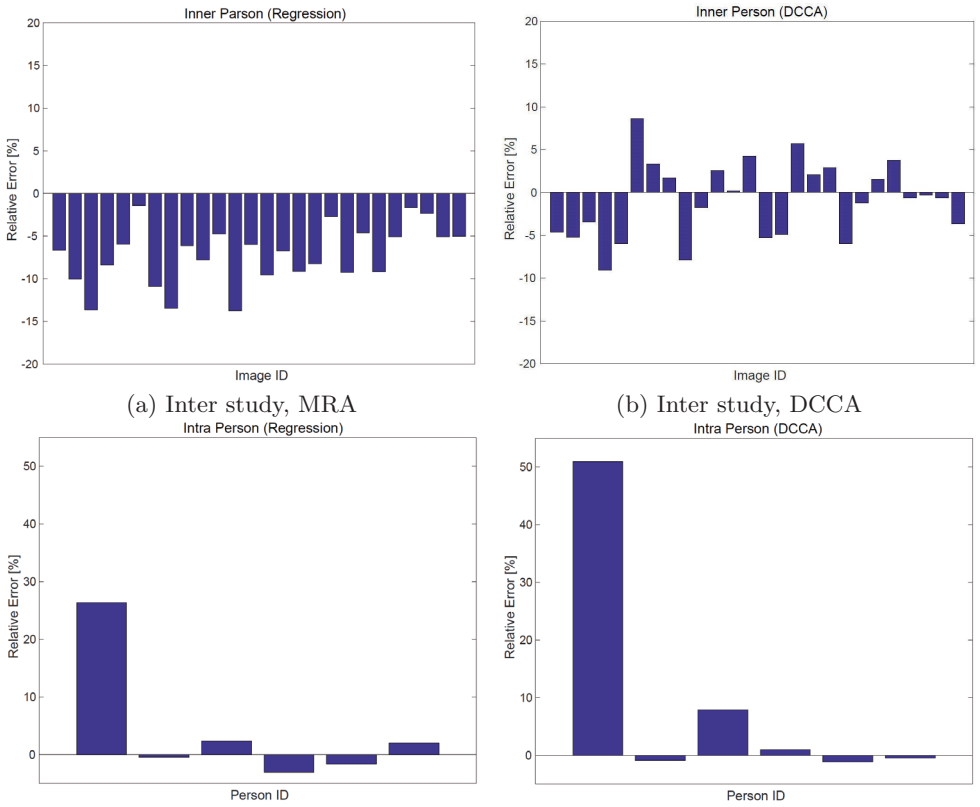
Twenty-six still stimulus images were displayed for 15 seconds each. Between two stimulus images, a white dot was displayed at the center of the monitor, and we asked the subjects to stare at it. The dot controlled the initial position for each image. Because the monitor reflected objects in the room, we kept the room dark so as to reduce reflections. We had the subjects freely see still stimulus images. Each stimulus image was shown for 15 seconds. We did not give any instructions on how the subject should see the images.

An eye tracker often misses eye direction. According to the specifications sheet, the eye tracker we used can correctly estimate $\pm 20^o$ horizontally and $\pm 40^o$ vertically. The range could not cover the whole of the monitor, then the eye tracker lost the eye direction when the subject saw the corners of the monitor. We ignored the eye tracking data including the missing frames both in learning and execution phases. Only the eye tracking data within the stimulus images was used for the regression model and the calculation of error. In our experimental environment, the limits were $\pm 29^o$ in horizontal and $\pm 17^o$ in vertical. We set no limit for the direction of head.

We adopted our proposed methods for eye tracking and head motion in time sequences. MRA and DCCA show the relationship of eye direction and head motion history of six subjects from the data for 25 stimulus images in the inter-user study as shown in **Figure 7(a)**. The estimation error was calculated by adopting the learning model into the data for one unused stimulus image. We validated how the learned model is generalized to images in this study. The situation of five subjects with 26 stimulus images in the intra-user study is shown in Figure 7(b). The error was calculated by adopting the learning model into the data of one unused subject. We validated how the learned model is generalized between subjects in this study. The design is called as leave-one-out cross validation. We set $\lambda_1 = 0.40$ and $\lambda_2 = 0.80$ for DCCA, $n = 5$ for the history frames, which were the best parameters through our trials and errors.

Figure 8 shows the reduction percentage of estimation error for each stimulus image in the inter-user and intra-user study. Figures 8(a) and 8(c) indicate the error reduction achieved by head motion history with MRA, and Figures 8(b) and 8(d) indicate the error reduction achieved by head motion history with DCCA. Figures 8(a) and 8(b) are the results of the inter-user study. Figures 8(a) and 8(b) show the median of estimation error for each image given by 6 subjects, and Figures 8(c) and 8(d) show the median of estimation error for each subject given by all images. Negative values mean that the accuracy of eye tracking was improved.

The average error for all stimulus images was -7.2% with MRA, and -0.8% with DCCA in the inter-user study. For most stimulus images and subjects, head motion history improved the accuracy of eye tracking. The absolute values of original estimation error were 29.1^o (S.D. 2.9^o) on average. By introducing the regression models, the



(a) Inter study, MRA (b) Inter study, DCCA

(c) Intra study, MRA (d) Intra study, DCCA

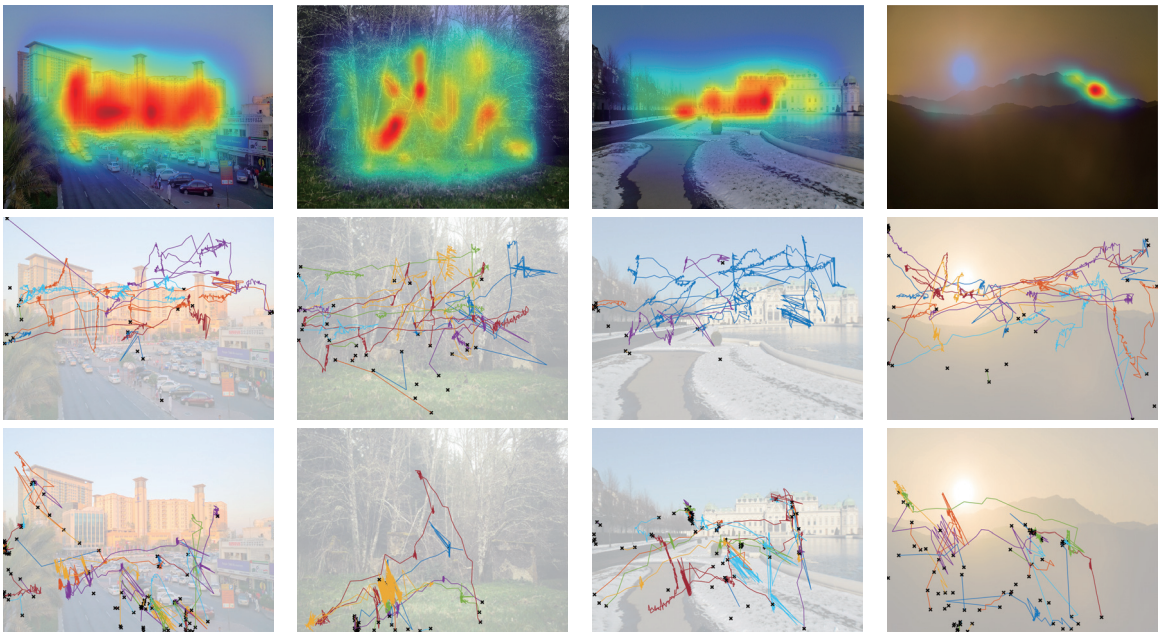
Fig. 8 Estimation errors of subject studies

error became 26.5^o (S.D. 2.4^o) on average in MRA and 29.2^o (S.D. 2.6^o) on average in DCCA. Compared with MRA of the simple model, DCCA did not improve the accuracy. According to the results of a one-sided t-test, only the result with MRA in the inter-study (Figure 8(a)) was significant ($p < 0.05$). There was no significant difference for the results with DCCA.

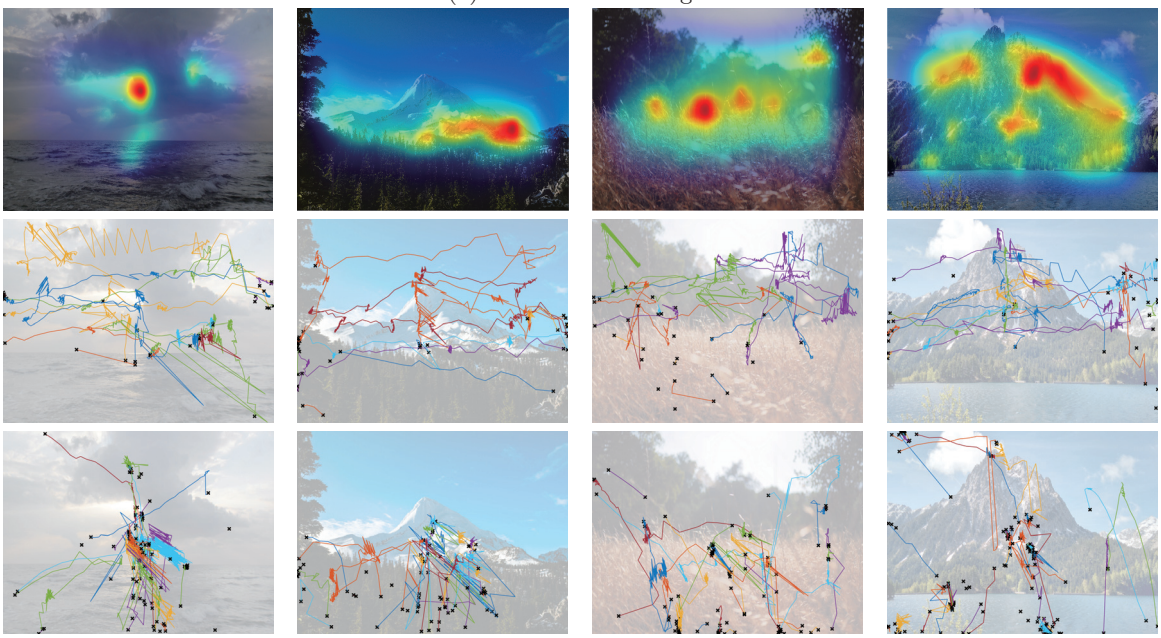
In the intra-user study, the average error was increased by 4.3% with MRA, and 9.6% with DCCA. The results for subject 1 were much worse than those for the other subjects, which degrades the total accuracy. The relationship of eye direction and head motion history in subject 1 could not be modeled well by the data of other subjects. Some private habits of viewing or poor eyesight of subject 1 could have affected the results.

We expected that DCCA would show the relationship of eye direction and head motion well, but the results of MRA were much better than those of DCCA. The eye movement containing high frequency components and the existence of missing or incorrect frames by eye blinking degrades the accuracy of estimation. It is possible that a simple model derived better results. In addition, it is clear that humans can change eye direction without moving their heads; therefore, some head motion is not related to eye direction, as we described in the introduction. As the result that we tried to estimate B and C with fixed S as the unit matrix, the error was almost same. Therefore, we concluded that the introduction of projection to low dimensional space could not contribute to estimate more accurate regression model.

There were large differences in results among the stimulus images. **Figure 9** shows the images that scored the best and the worst for improving the accuracy of eye tracking. We generated saliency map with a graph-based visual saliency (GBVS) algorithm[9]; and overlaid it on the stimulus images. The best images have the tendency to include batch of areas with high saliency values, whereas the worst images include cluttered small areas. From the eye tracking data, the worse images tend to attract the gaze of subjects in not salient regions, and individual subjects saw different regions of the images. The best images attracted the gaze of many subjects in similar way and in similar regions, which would contribute more accurate estimation of eye direction. However, there are exceptions within the images representing eye tracking data. The tendency would differ from much more subjects. We would like to examine how much the accuracy can be improved with much more data as future work. In this work, we at least confirmed that it is possible to improve the accuracy with head motion history, and MRA is more



(a) Best stimulus images



(b) Worst stimulus images

Fig. 9 The best and worst stimulus images for improving the accuracy. (Top) Saliency maps for stimulus images, (middle) eye tracking data of subject 1, (bottom) eye tracking data of subject 4. The cross marks (x) in the figure represent that we failed to track the eye direction in its previous or following frames because the subject saw outside of the image, or blinked his/her eyes. We confirmed that eye tracking data includes high frequency components and we failed to track the eye direction in many frames.

promising than DCCA for the purpose.

Saccades do not involve the movement of head, therefore images with distributed saliency points are difficult gives worse results in the experiment. This is the limitation of our method. Our method would work well in the case that saccades are not observed.

5. Discussions and Future Work

We proposed a method for improving eye direction estimation by head motion history. In the experiments, the accuracy of eye tracking was improved with MRA and DCCA models. As an important future work, we will analyze the results from the view of subjects and stimulus images for specifying the valid situations. In many applications, discrete-time eye tracking data may be more

useful than constant eye tracking data.

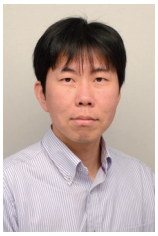
Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers JP17H00737, JP16K12784, JP17H00738.

References

1) EMR-9, NAC Image Technology.
2) K. Sankaranarayanan, M. C. Chang, N. Krahnstoever: "Tracking gaze direction from far-field surveillance cameras," Proc. Winter Conference on Applications of Computer Vision, pp.519–526 (2011).
3) C. Lee: "Eye and head coordination in reading: roles of head movement and cognitive control," Vision Research, Vol. 39, pp.3761–3768 (1999).
4) Y. Okinaka, I. Mitsugami, Y. Yagi: "Gaze Estimation Based on Eyeball-Head Dynamics," Proc. IPSJ CVIM, Vol. 2016-CVIM-202, Article 18 (2016).
5) M. Toyoura, T. Tanaka, A. Sugiura, X. Mao: "Eye Tracking by Head Motion History," Proc. International Workshop on Image Electronics and Visual Computing, pp.1–5, Article 4B-3 (2017).
6) W. Einhüser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, P. König: "Human eye-head coordination in natural exploration," Network: Computation in Neural Systems, pp.267–297 (2007).
7) F. De la Torre, M. J. Black: "Dynamic coupled component analysis," Proc. CVPR, Vol. 2, pp.II-643-II-650 (2001).
8) X. Ma, Z. Deng: "Natural eye motion synthesis by modeling gaze-head coupling," Proc. IEEE VR, pp.143–150 (2009).
9) J. Harel, C. Koch, P. Perona: "Graph-based visual saliency," Advances in Neural Information Processing Systems, pp.545–552 (2006).

(Received July 17, 2017)
(Revised November 10, 2017)



Masahiro TOYOURA (*Member*)
He received the B.Sc. degree in Engineering, M.Sc. and Ph.D. degrees in Informatics from Kyoto University in 2003, 2005 and 2008 respectively. He is currently an Associate Professor at Interdisciplinary Graduate School, University of Yamanashi, Japan. His research interests are augmented reality, computer and human vision. He is a member of ACM and IEEE Computer Society.

Takumi TANAKA
He received the B.Sc. degree in Engineering from University of Yamanashi, Japan. His research interests include computer vision and human sensing.

Atsushi SUGIURA
He received the B.Sc. degree in Engineering, M.Sc. and Ph.D. degrees in Engineering from University of Yamanashi in 2002, 2012 and 2015 respectively. He is currently an Assistant Professor at Interdisciplinary Graduate School, University of Yamanashi, Japan. His research interests are augmented reality, computer and human vision.

Xiaoyang MAO (*Member*)
She received her B.Sc. in Computer Science from Fudan University, China, M.Sc. and Ph.D. in Computer Science from University of Tokyo. She is currently a Professor at Interdisciplinary Graduate School, University of Yamanashi, Japan. Her research interests include texture synthesis, non-photo-realistic rendering and their application to scientific visualization. She is a member of ACM SIGRAPH and IEEE Computer Society.