EYE TRACKING BY HEAD MOTION HISTORY

Masahiro Toyoura Takumi Tanaka Atsushi Sugiura Xiaoyang Mao
University of Yamanashi

ABSTRACT

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 data regarding human attention by surveillance camera or an intuitive interface for tablet devices. Through a preliminary experiment, we found that head direction is often largely different from eye direction. We propose to estimate more 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.

1. INTRODUCTION

Eye direction is a promising cue for understanding a user’s attention. Surveillance camera images can detect human attention for public signs by eye tracking. The most common technique for estimating eye direction is capturing images of human eyes [1]. Cameras are installed on a person’s glasses or the frontal side of a monitor display. 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 do not wear glasses or other special devices. However, tracking eye movement with head motion is not as accurate as tracking eye movement with cameras. Sankaranarayanan et. al. proposed a method for eye tracking of pedestrians with 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 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 completely. 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.

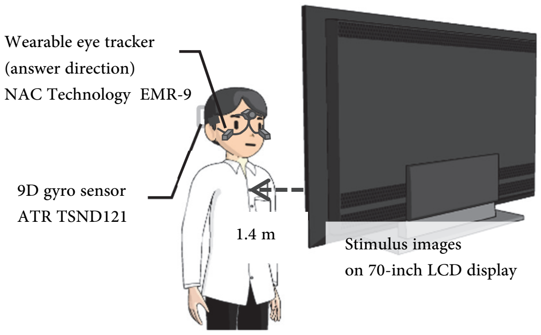


Figure 1. Environment of experiment.

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 situations 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, and we also discuss the difference in accuracy between stimulus images.

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 can estimate 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 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 given moment. We used multiple regression analysis (MRA) and dynamic coupled component analysis (DCCA) [5] 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 which substantively controls the input and output spaces, and shows the projections of input and output vectors for the lower dimensional space. DCCA works well even if 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 [6] 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. The only disadvantage is the lower speed of convergence because of the complexity of the model.

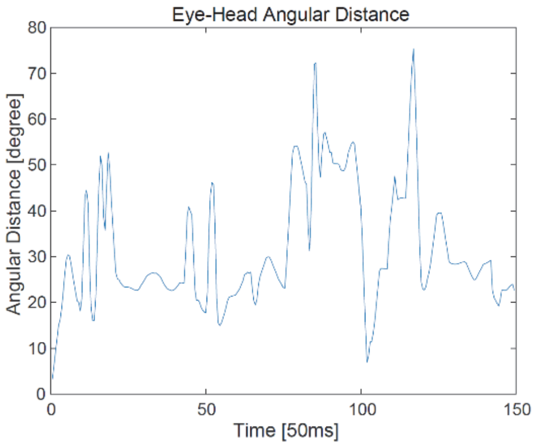
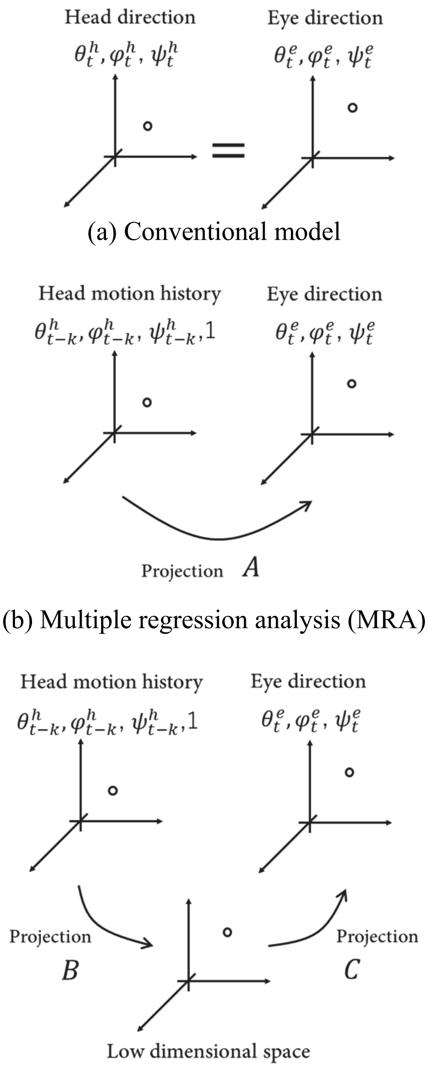


Figure 2. Error of eye tracking by head motion.


Head direction
$\theta_t^h, \varphi_t^h, \psi_t^h$
Eye direction
$\theta_t^e, \varphi_t^e, \psi_t^e$
=

(a) Conventional model

Head motion history
$\theta_{t-k}^h, \varphi_{t-k}^h, \psi_{t-k}^h, 1$
Eye direction
$\theta_t^e, \varphi_t^e, \psi_t^e$
Projection A

(b) Multiple regression analysis (MRA)

Head motion history
$\theta_{t-k}^h, \varphi_{t-k}^h, \psi_{t-k}^h, 1$
Eye direction
$\theta_t^e, \varphi_t^e, \psi_t^e$
Projection
B
Projection
C
Low dimensional space

(c) Dynamic coupled component analysis (DCCA)

Figure 3. Differences among three learning models.

Figure 3 shows the difference among three types of learning. The input vector is roll θ_{t-k}^h, pitch φ_{t-k}^h and yaw ψ_{t-k}^h of head direction at time t, its previous time $t-k$ $(k=0,\cdots,n)$ and constant 1. n is set to 5 in the experiments. Total dimension is $3(n+1)+1$. The output vector is roll θ_t^e, pitch φ_t^e, and yaw ψ_t^e of eye direction at time t. The dimension is 3. In the conventional model shown in Figure 3(a), eye direction is estimated as the same direction of the head. MRA, in Figure 3(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 3(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.

$$\mathrm{e}_t = \begin{pmatrix} \theta_t^e \\ \varphi_t^e \\ \psi_t^e \end{pmatrix}, h_t = \begin{pmatrix} \theta_{t-n}^h \\ \varphi_{t-n}^h \\ \psi_{t-n}^h \\ \vdots \\ \theta_t^h \\ \varphi_t^h \\ \psi_t^h \\ 1 \end{pmatrix},$$

$$E = \sum^{t} \| e_h - A h_t \|. \tag{1}$$

DCCA treats the projection matrices B and C as follows. S is a temporal smoothing matrix. The hat notations mean that the dimensions are reduced by PCA or other techniques. λ_1 and λ_2 are arbitrary constants for weighting the terms.

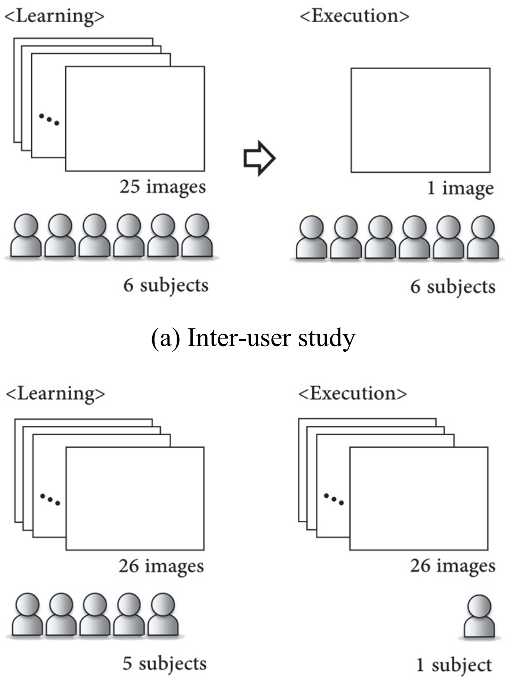
$$E_{static} = \sum^{t} \left(\| \hat{h}_t - C e_t \|_2^2 + \lambda_1 \| e_t - B h_t \|_2^2 \right), \tag{2}$$

$$E_{dynamic} = E_{static} + \lambda_2 \sum^{t} \| e_t - S e_{t-1} \|_2^2. \tag{3}$$

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.

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. They 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.



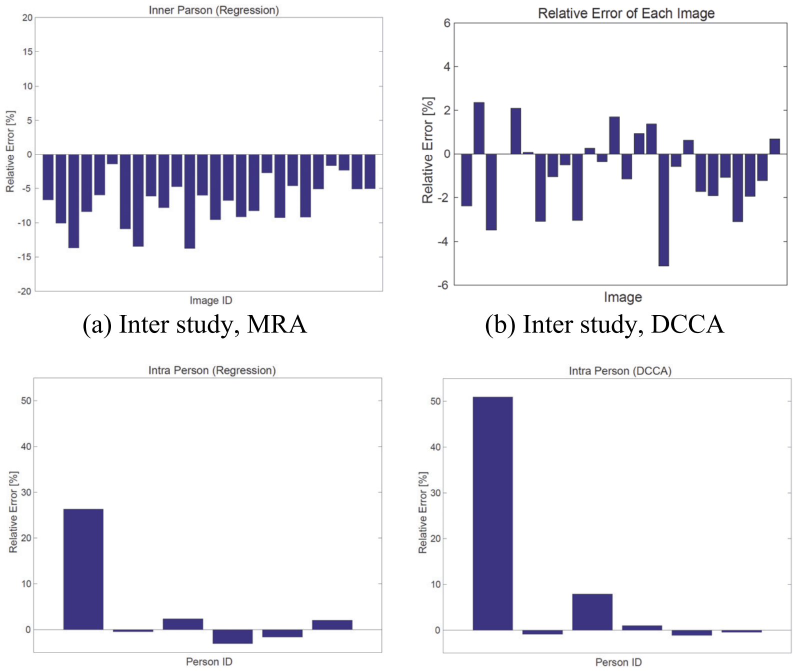
(a) Inter-user study

(b) Intra-user study

Figure 4. Design of subject studies.

An eye tracker often misses eye direction. According to the specifications sheet, the eye tracker we used can correctly estimate ±20° horizontally and ±40° 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.

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 shown in Figure 4(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 data of five subjects with 26 stimulus images in the intra-user study are shown in Figure 4(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 to 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.



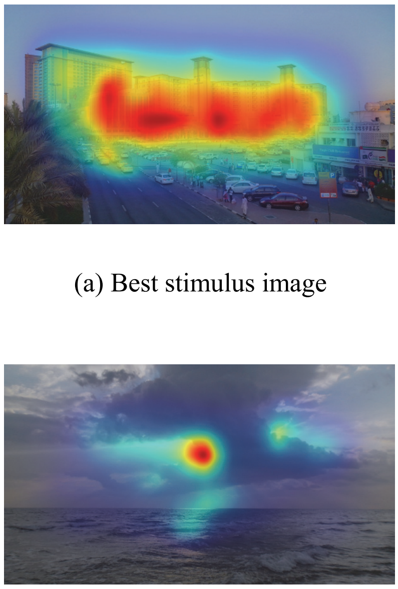
(a) Inter study, MRA

(b) Inter study, DCCA

(c) Intra study, MRA

(d) Intra study, DCCA

Figure 5. Estimation errors of subject studies.



(a) Best stimulus image

(a) Worst stimulus image

Figure 6. The best and worst stimulus images for improving the accuracy.

Figure 5 shows the reduction percentage of estimation error for each stimulus image in the inter-user study. Figures 5(a) and 5(c) indicate the error reduction achieved by head motion history with MRA, and Figures 5(b) and 5(d) indicate the error reduction achieved by head motion history with DCCA. Figures 5(a) and 5(b) are the results of the inter-user study. The values are the median for each image/subject of the median of subjects/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. For most stimulus images and subjects, head motion history improved the accuracy of eye tracking. 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 5(a)) was significant ($p<0.05$).

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. Because the obtained data were essentially noisy data, it is possible that a simple model derived better results. As we described in the introduction, it is clear that humans can change eye direction without moving their heads; therefore, some head motion is not related to eye direction.

There were large differences in results among the stimulus images. Figure 6 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 [7]; and overlaid it on the stimulus images. The best image had a large area with high saliency values, whereas the worst image had several small areas with high saliency values. We could see the tendencies for the other stimulus images; i.e., scattered salient areas evoked errors of eye tracking with our method. Saccade, or fast eyeball movement, does not involve head motion, so eye direction could not be estimated from head motion. Our method would work well in the case that saccades are not observed.

5. DISCUSSIONS AND FUTURE WORKS

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.

6. ACKNOWLEGEMENT

This work was partially supported by JSPS Grants-in-Aid for Scientific Research KAKENHI (JP26240015).

7. REFERENCES

[1] NAC Technology, EMR-9.

[2] K. Sankaranarayanan, M. C. Chang, and N. Krahnstoever, "Tracking gaze direction from far-field surveillance cameras," WACV, 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, and Y. Yagi, "Gaze Estimation Based on Eyeball-Head Dynamics," vol. 2016-CVIM-202, Article 18, 2016.

[5] F. De la Torre and M. J. Black, "Dynamic coupled component analysis," in Computer Vision and Pattern Recognition, CVPR, pp. II-643-II-650 vol. 2, 2001.

[6] X. Ma and Z. Deng, "Natural eye motion synthesis by modeling gaze-head coupling," IEEE VR, pp. 143-150, 2009.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," Advances in Neural Information Processing Systems, pp. 545-552, 2006.