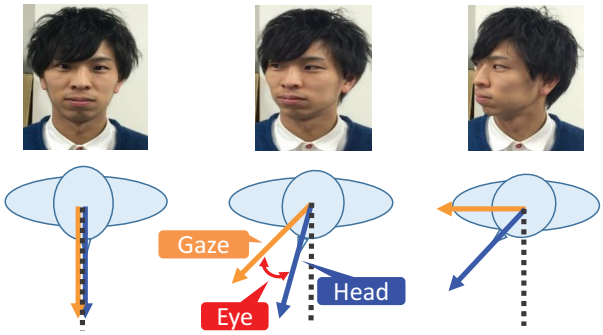
Gaze Estimation Based on Eyeball-Head Dynamics

Ikuhisa Mitsugami Yamato Okinaka Yasushi Yagi
Osaka University
{mitsugami,okinaka,yagi}@am.sanken.osaka-u.ac.jp

Abstract

Human's gaze direction is a useful cue to understand his/her attention and interest. There are many kinds of eye tracking devices, but they are usually unavailable for people observed in surveillance views because they are located too far to observe their eyeballs. If we could know the gaze direction from such surveillance views, it should be very effective for many applications. Our research objective is thus to estimate the gaze direction without any eye-trackers but by observing their behaviors. This paper proposes a new gaze estimation method based on a dynamical model emulating dynamic relation between eyeball and head. Experimental results using head-mounted and body-mounted camera images confirmed its effectiveness.



(a) 0 deg (b) 45 deg (c) 90 deg

Figure 1. Head direction is not identical to gaze direction.

1. Introduction

When a person is interested in a certain object, he/she naturally gazes at the object. Gaze direction is thus a useful cue to understand his/her attention and interest. If we can know gaze directions of people who appear in surveillance camera images, it would be useful for many application; for example, we could recommend items to a person who looks interested in them, and we could predict a shoplifter who gaze at not goods but security cameras or store clerks.

For obtaining gaze direction, eye-trackers are the most popular way. There are two types of eye-trackers; a wearable and stationary ones. In many studies use the wearable one for measuring the gaze of a person who moves freely [1, 2]. On the other hand, the stationary ones are used for a display [3]. These methods are, of course, give correct gaze direction. They are, however, not suitable for the applications above mentioned; a shoplifter will never wear the eye-tracker, and the stationary ones are useless for people in a wide area. Recently, there are several studies that estimate gaze direction by cameras, but even by the state-of-the-art method such as [4] a person has to appear quite large in captured images. It is usually impossible to estimate gaze from surveillance views because they are located too far to observe their eyeballs.

Another possible way is to use head directions instead of the gaze directions. Even in surveillance views, heads appear much large than eyeballs. Thanks to development of face detection and face direction estimation, it is possible to obtain the head directions in most scenes. In fact, in many studies related to attention estimation in surveillance views, this approach are often used [5].

Considering our purpose and the summary of related works above mentioned, it is reasonable to choose the camera-based way that rely on that the head direction can be good approximation of the gaze direction. However, we doubt this approximation considering the following observation. Fig. 1 shows human faces naturally directing to three directions; (a) 0 degree (to the front), (b) 45 degree and (c) 90 degree (the side). As shown in this figure, a person look to a certain direction by combining rotations of head and eyeballs, so that he/she never look at the direction only by the rotation, as reported in [6, 7, 8]. Moreover, this figure is just about static situations, but when he/she change his/her attention dynamically, the pose of head and eyeballs show more complex cooperative motions, which is known as Vestibulo-Ocular Reflex (VOR). Considering this observation, this approximation is not correct so that we should not estimate the gaze direction relying on it.

Aiming at estimating the gaze direction without any

978-1-5090-4941-7/17 $31.00 © 2017 IEEE
DOI 10.1109/WACVW.2017.15

eye-trackers, therefore, we proposes a new gaze estimation method by modeling the static and dynamic cooperation of head and eyeballs. In this method, we approximate this head-eye cooperative model by a simple dynamic model. Thanks to this approximation, the model can be formulated as a differential equation. This equation then can be solved linearly by sequences of subject's eyeballs, head and chest, which is needed to define front for each moment. We evaluate performance of this proposed method by comparing the result of our proposed method and the method which regards head direction as gaze direction with measured gaze direction. In this experiment, we use head-mounted and body-mounted cameras to accurately and easily obtain head and chest directions by Structure-from-Motion (SfM). We finally confirm effectiveness of the proposed method.

It should be noted that the proposed method is not only for egocentric videos. The reason why we used them in our experiments is just to easily obtain poses of the head and chest. It would be possible to apply our method even for surveillance videos if we are equipped with precise head/chest pose estimation method, which is out of scope of this paper. In fact, if we limit to the egocentric video scenario, there are several studies [12, 13, 14].

2. Eye-head coordination

This section describes the eye-head coordination of human [9, 10]. When a person shifts his/her gaze direction, he/she moves his/her head as well as eyeballs. According to [6, 7, 8], when he/she keeps gazing at a certain direction, rotation angles of the head and eyeballs have linear relation. In addition, when he/she changes his/her gaze to another direction, which is called "saccade," the head and eyeballs show characteristic cooperation as shown in Fig. 2. In the early phase of a shift, a eyeball moves to a next direction rapidly. This motion, i.e. saccade, occurs in order to catch the target in the center of the retina. On the other hand, the head also moves in the same direction, but starts moving a little later, because the head has much larger mass than the eyeball and so is hard to move quickly. After the head start moving, the eyeball begin to move to the opposite direction so as to be symmetry with the head motion. This is considered for make images on retina stable. This neural control of head and eyeballs is called vestibulo-ocular reflex (VOR).

Let us see Fig. 2 again. As a result of VOR, the gaze and head show the following behaviors: When the gaze shift occurs, the head starts following the gaze at slower speed than the gaze and converges to a certain angle after enough length of time. This relation is in fact often observed in real data. Fig. 3 shows an example of the gaze and head motion we collected. We can see many pairs of the gaze and head motion that follow this rule.

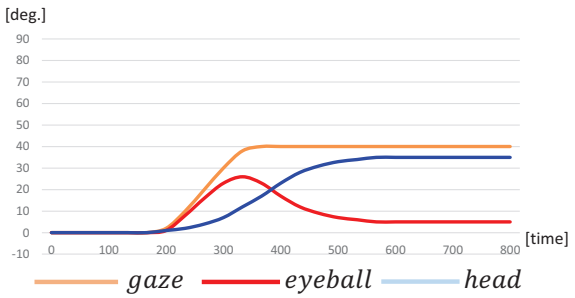


Figure 2. Eye-head coordination

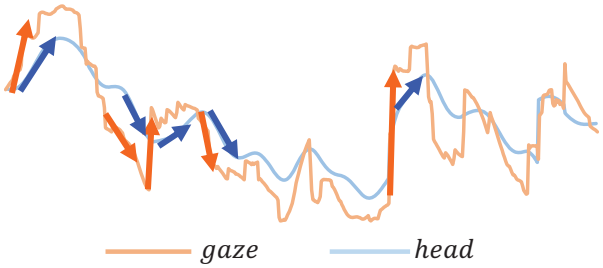


Figure 3. Eye-head coordination observed in real human behavior.

3. Method

3.1. Approximation Model

Let us imagine two balls connected to each other by a spring as shown in Fig. 4. We define that a yellow ball can move freely, and its mass and force the ball receive are ignored. We also define, on the other hand, a blue ball passively moves pulled by the gaze ball via the spring. Moreover, there is a damper for the blue ball, so that when it is pulled by the yellow one it can move just gradually (Fig. 5). In this setting, when the yellow ball is suddenly moved in a certain length, the blue ball would then move afterwards with the smaller speed, then finally stop. This situation can be formulated as follows:

$$\begin{aligned} F = mh''(t) &= k\{g(t) - h(t) - l\} - \lambda h'(t) \\ \iff g(t) &= ah(t) + bh'(t) + ch''(t) + d \end{aligned} \quad (1)$$

where m denotes mass of the blue ball, l denotes natural length of the spring, and $g(t), h(t)$ denote motions of the yellow and blue balls, respectively.

Considering the observation and discussion in Section 2, the eye-head coordination looks very similar to the behaviors of the balls. Conversely, the eye-head coordination should be well approximated by the formulation for them. Based on this discussion, we adopt Equation 1 as an approximation model of the eye-head coordination.

3.2. Gaze Estimation Method

It is apparent that once we obtain the model parameters a, b, c, d we should just input observed head motion to

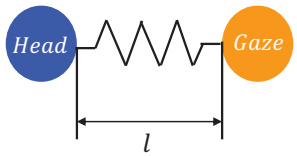


Figure 4. A dynamic model of two ball connected to each other by a spring.

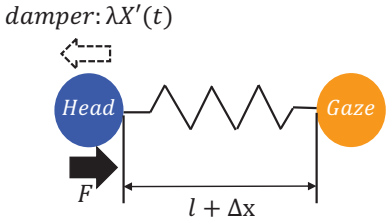


Figure 5. Dynamics when the yellow ball moves aprt from the blue one.

Equation 1 to obtain estimate of the gaze motion. To obtain the parameters, we collect multiple pairs of the gaze and head motion beforehand, and assign them to the following linear system:

$$\begin{bmatrix} g_1 \\ \vdots \\ g_T \end{bmatrix} = \begin{bmatrix} h_1 & h_1' & h_1'' & 1 \\ \vdots & \vdots & \vdots & \vdots \\ h_T & h_T' & h_T'' & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \tag{2}$$

where h_t, g_t denote values of $h(t), g(t)$ at time t. ($h(t) = [h_1, h_2, \cdots, h_t]$, $g(t) = [g_1, g_2, \cdots, g_t]$.) By multiplying pseudo inverse matrix from the left, we can obtain a, b, c, d. Note that we apply RANSAC for this linear sysmtem to make the method robust to noises, which are often included in the gaze motions.

4. Experiment

4.1. Experimental setting

To collect a person's natural behavior while maintaining high accuracy of motion measurement, we adopted wearable-camera-based motion capture [11]. each participant wore wearable cameras (GoPro HERO3) on his/her chest and head so as to observe his/her front, as shown in Fig. 6. In addition, he/she also wore a wearable eye-tracker (NAC EMR-9) to obtain true gaze history. We applied SfM to all captured data to reconstruct all 3-D motions of the cameras (including a camera equipped in the eye-tracker) as well as 3-D structure of environment.

In the experiments, there were eight participants, but in this paper we picked up three of them because the others' motion data looked less reliable due to bad reconstruction by SfM. The model parameter estimation was performed for each person, since we considered that each person should have different gazing manner.

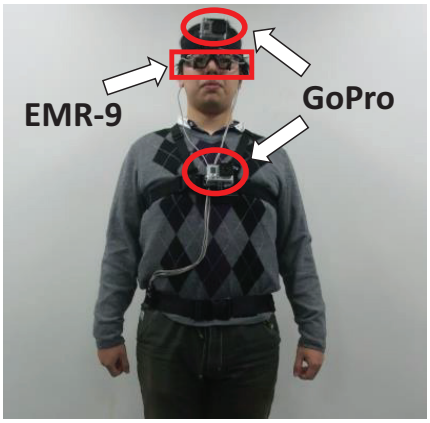


Figure 6. Experimental setting for a participant.

Table 1. Mean absolute error in horizontal angles.

	Simple	Proposed
MAE	11.6	7.9

Note that we define the front of a participant as his/her chest direction . Thus relative angles of the gaze and head to the chest were used as $g(t), h(t)$, respectively.

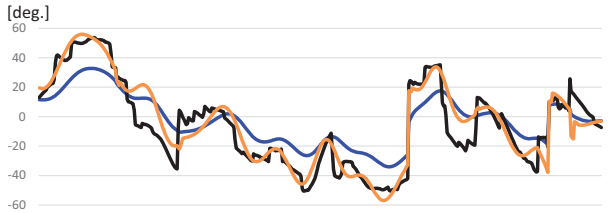
4.2. Evaluation of Model Validity

In this section, to validate this model approximation, we used the same sequences of the gaze and head for calculating the model parameters and as the testing data. More realistic performance evaluation using different sequences for training and testing are described in the next section.

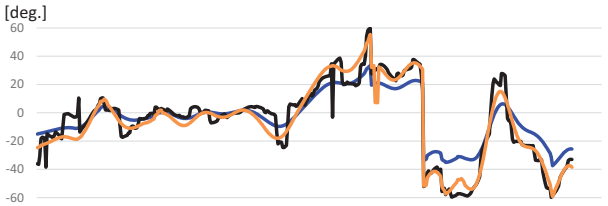
4.2.1 Horizontal Direction

Fig. 7 shows experimental results for the three participants. Black sequences denote the ground truth captured by the eye-tracker, and orange ones denote $g(t)$ estimated by our proposed method. For comparison, there are also blue sequences describing the head direction. Though it is indeed a test input for the proposed method, it also has another meaning; another estimate of the gaze based on the assumption that the head direction well approximates the gaze ones. We call it "simple method" afterwards.

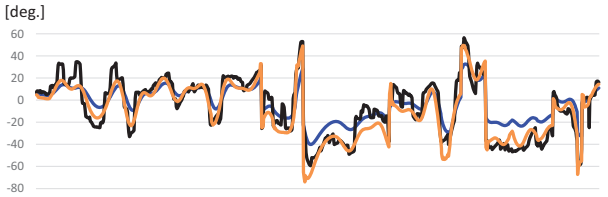
From these graphs, we confirm that the estimates by the proposed method apparently outperforms those by the simple method. In addition, quantitative evaluation is shown in Fig. 8, which are error histograms of the simple and proposed methods. Table 1 shows averages of these histograms. From the figure and table, the effectiveness of the proposed method are re-confirmed.



(a) Participant A.

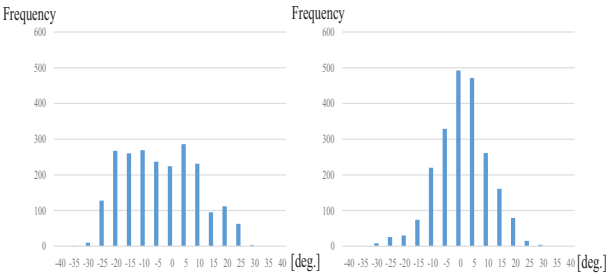


(b) Participant B.



(c) Participant C.

Figure 7. Model regression for horizontal angles.



(a) Simple (b) Proposed

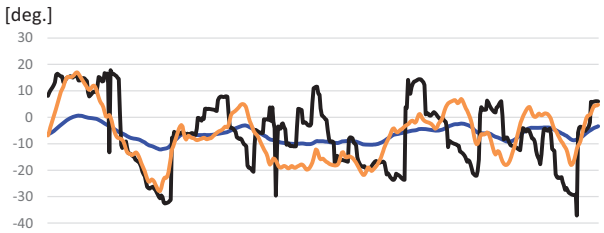
Figure 8. Error histograms for horizontal angles.

Table 2. Mean absolute error in horizontal angles.

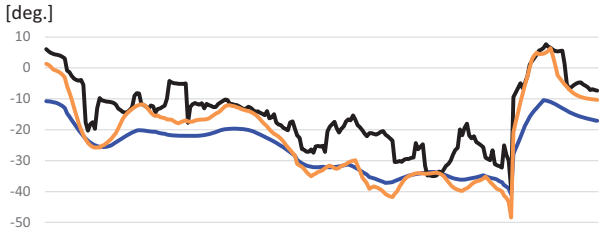
	Simple	Proposed
MAE	7.9	6.8

4.2.2 Vertical direction

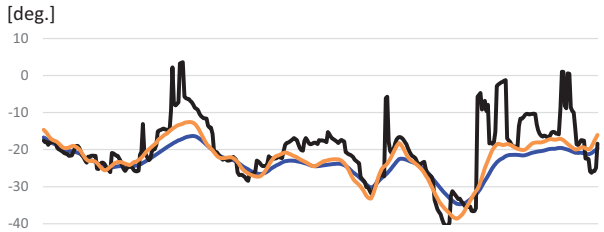
We also performed the same experiments for vertical angles Fig. 9, Fig. 10, and Table 2 show its results. As a result, we did not confirm effectiveness of the proposed method; it just gave similar accuracy to the simple method. We consider this is because of the shape of eye; wide horizontal range while narrow vertical range.



(a) Participant A.

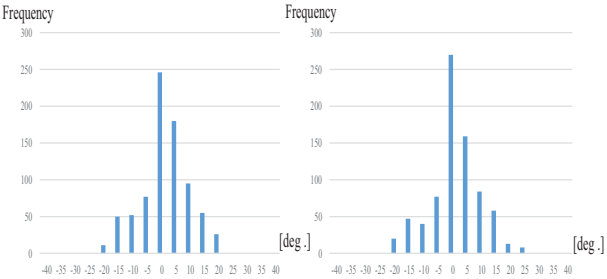


(b) Participant B.



(c) Participant C.

Figure 9. Model regression for horizontal angles.



(a) Simple (b) Proposed

Figure 10. Error histograms for horizontal angles.

4.3. Estimation result

For evaluating performance of the proposed method in realistic situation, we applied cross validation using two different sequences for each participant. Table 3 summarizes the results.

5. Conclusion

In this paper, we proposed a new gaze estimation method by modeling the static and dynamic cooperation of head and eyeballs. We adopted an approximation model that well em-

Table 3. Results of cross validation.

Participant	Training	Test	Simple	Proposed
A	scene1	scene2	9.2	7.7
	scene2	scene1	9.7	8.2
B	scene3	scene4	22.2	16.8
	scene4	scene3	14.9	9.9
C	scene5	scene6	15.6	10.9
	scene6	scene5	12.9	9.5

ulate the head-eye cooperative model. Since the model can be formulated as a differential equation, it can be solved linearly by sequences of subject's eyeballs and head. We evaluate performance of this proposed method by comparing the result of our proposed method and the method which regards head direction as gaze direction with measured gaze direction. We finally confirm effectiveness of the proposed method by applying it to the gaze and head motions collected from real participants. The gaze and head motion were measured by putting eye-trackers and cameras on their head and chest and applying SfM to all images captured by them.

Since this study is still on an early stage, there are many future works remaining. One of the important tasks is investigation about variation of the model parameters among different subjects. If they are similar, the model parameters can be shared for all people, thus we do not need to worry about calibration for each subject. If, however, the model parameters are unique for each person, we need to propose the calibration method.

Another problem is about reasonability of the model itself. In the current model, control of the gaze and head are defined by a spring. However, it implicitly contains repelling forces when these are nearer than it natural length, which should never occur in the eye-head coordination. We might need to deeply consider and design the improved model.

Application to surveillance views is also an important topic. In this paper, in order to prepare accurate motion data, we use wearable devices. Considering the motivation described in Section 1, however, we should cope with surveillance views without any wearable devices. We will realize that by combining the proposed method and state-of-the-art human pose estimation techniques.

References

[1] I. Mitsugami, N. Ukita, M. Kidode, "Estimation of 3D Gazed Position Using View Lines," 12th International Conference on Image Analysis and Processing, 2003.

[2] V. Rantanen, T. Vanhala, O. Tuisku, P. H. Niemenlehto, J. Verho, V. Surakka, M. Juhola, J. Lekkala, "A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction," IEEE Transactions on Information Technology in Biomedicine, pp.795–801, 2011.

[3] H. Konno, T. Gotoh, T. Takegami, "Method for Multidimensional Operation Interface Using Eye Location Detection," The Journal of The Institute of Image Information and Television Engineers, pp. 518–525, 2007.

[4] F. Lu, Y. Sugano, T. Okabe, Y. Sato, "Adaptive Linear Regression for Appearance-Based Gaze Estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.36, No.10, pp.2033–2046,2014.

[5] S. O. Ba, J. M. Odobez, "Recognizing visual focus of attention from head pose in natural meetings," IEEE Transactions on Systems, Man, and Cybernetics, Part B:Cyberneticspp, pp.1886–1893, 2009.

[6] T. Okada, H. Yamazoe, I. Mitsugami, Y. Yagi, "Preliminaly analysis of gait changes that correspond to gaze direcions," 2nd IAPR Asian Conference on Pattern Recognition, pp.788–792, 2013. pp. 788–792, Nov. 2013.

[7] Y. Fang, M. Emoto, R. Nakashima, K. Matsumiya, I. Kuriki, S. Shioiri, "Eye-position distribution depending on head orientation when observing movies on ultrahigh-definition television," ITE Transactions on Media Technology and Applications, Vol.3, No.2, pp.149–154, 2015.

[8] Y. Fang, R. Nakashima, K. Matsumiya, I. Kuriki, S. Shioiri, "Eye-head coordination for visual cognitive processing," PLoS ONE, Vol.10, No.3, e0121035, 2015.

[9] T Maesako, T. Koike, Measurement of coordination of eye and head movements by sensor of terrestrial magnetism Japanese Journal of Physiological Psychology and Psychophysiology Vol.11, No.2 pp.69–76, 1993.

[10] G. M. Jones, D. Guitton, A. Berthoz, "Changing patterns of eye-head coordination during 6 h of optically reversed vision," Experimental Brain Research, Vol.69, No.3, pp.531–544, 1988.

[11] T. Shiratori, H. S. Park, L. Sigal, Y. Sheikh, J. K. Hodgins "Motion Capture from Body-Mounted Cameras," ACM Transactions on Graphics, Vol.30, No.4, 2011.

[12] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto, K. Hiraki, "Attention prediction in egocentric video using motion and visual saliency," Pacific-Rim Symposium on Image and Video Technology, pp.277-288, 2012.

[13] Y. Li, A. Fathi, J. M. Rehg, "Learning to predict gaze in egocentric video," International Conference on Computer Vision, pp.3216–3223, 2013.

[14] T. Leelasawassuk, D. Damen, W. W. Mayol-Cuevas, "Estimating Visual Attention from a Head Mounted IMU," International Symposium on Wearable Computers, pp.147–150, 2015.