Gaze Estimation Using Regression Analysis and AAMs Parameters Selected Based on Information Criterion

M.Takatani, Yasuo Ariki, and Tetsuya Takiguchi

Graduate School of Engineering, Kobe University,
Organization of Advanced Science and Technology, Kobe University
takatani_m@ me.cs.scitec.kobe-u.ac.jp
{ariki, takigu}@kobe-u.ac.jp
http://www.me.cs.scitec.kobe-u.ac.jp/

Abstract. One of the most crucial techniques associated with Computer Vision is technology that deals with the automatic estimation of gaze orientation. In this paper, a method is proposed to estimate horizontal gaze orientation from a monocular camera image using the parameters of Active Appearance Models (AAM) selected based on several model selection methods. The proposed method can estimate horizontal gaze orientation more precisely than the conventional method (Ishikawa's method) because of the following two unique points: simultaneous estimation of horizontal head pose and gaze orientation, and the most suitable model formula for regression selected based on each model selection method. The validity of the proposed method was confirmed by experimental results.

1 Introduction

The human gaze is thought to be effective for understanding or measuring the degree of his / her interest or attention because the information from the gaze is the most vital for humans to understand their environment. Thus, estimating gaze orientation automatically is expected to be applied not only to robot vision, artificial intelligence, and human interaction but also to the analysis of image and video content, and analysis and retrieval based on human percipient models [1, 2].

In order to estimate gaze orientation, two main types of methods have been proposed. One approach employs a special device (such as an infrared camera) as proposed by Ohno [3]. This approach can estimate gaze orientation with a high degree of accuracy. The other approach processes monocular camera images. The advantage of this approach is that gaze orientation can be estimated inexpensively because only a monocular camera is required. From this view point, we employ the latter approach in this study.

Many methods have been proposed for estimating gaze orientation from a monocular image. For instance, Yamazoe proposed the use of the Lukas-Kanade's feature tracking method [4] and 3D-eyeball model[5]. Gaze can be estimated stably by this method, even if the subject is not included in the training data. However, the precision is not so accurate because the gaze orientation is estimated after the head pose estimation.

Ishikawa proposed the use of 3D AAM (Active Appearance Models) to extract the coordinates of the feature points, and the gaze orientation was estimated by the 3D eyeball model. The gaze orientation can be computed more precisely than Yamazoe's method due to the improvement of head pose estimation error using this method. However, the positioning error of the feature points causes the gaze estimation error because gaze orientation is computed using the coordinates of the feature points relative to the eye [6].

Thus, there were few methods that address the relationship between head pose and gaze orientation simultaneously. This is the reason why we propose a method in this paper to estimate them simultaneously using regression-based AAM parameters.

Moreover, the feature parameters extracted by AAM contain unessential information for the estimation. To select the essential feature parameters, we employ the model selection method (e.g. AIC[10], MDL[11], BIC[12]).

The rest of this paper is organized as follows. In Section 2, the method to estimate horizontal gaze orientation is proposed. Experimental results are presented in Section 3, followed by concluding remarks in Section 4.

2 Proposed Method

In this section, the method to estimate horizontal gaze orientation is proposed. Fig. 1 shows a processing flow of the proposed method. First, the facial area in the test image is detected using AdaBoost based on Haar-like features for stable AAM performance. Next, the feature parameters are extracted by AAM on this facial area. Finally, the head pose and the horizontal gaze orientation are simultaneously estimated using a regression model that is selected based on AIC (Akaike Information Criterion), MDL (Minimum Description Length), and BIC (Bayes Information Criterion).

2.1 Facial Area Search

The performance of AAM feature extraction depends on the initial search points. To make AAM search performance more stable, the facial area in an image is roughly computed using AdaBoost based on Haar-like features proposed by Viola [7]. Haar-like features for face detection are based on the difference between the sums of the pixels within two rectangular regions of the same size and shape that are adjacent to one another horizontally or vertically.

Since the total number of Haar-like features is far larger than the number of pixels on the image, simple and efficient classifiers can be constructed by selecting a small number of important features using AdaBoost from a huge library of potential features.

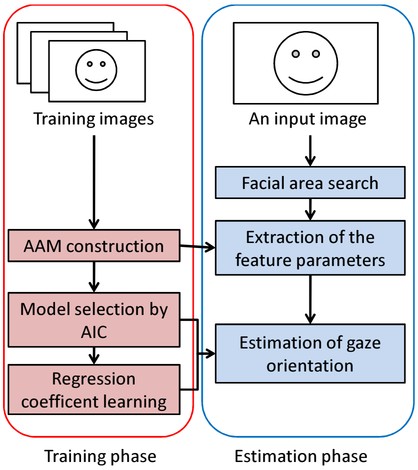

Fig. 1. Processing flow of the proposed method.

Actually, we employed "haarcascase_frontalface″ in OpenCV library for searching facial area.

2.2 Active Appearance Models

Cootes proposed AAM to represent shape and texture variations of an object with a low dimensional parameter vector $\mathbf{c}$ [8]. Vector $\mathbf{c}$ can represent various facial images with arbitrary orientation of face and gaze using the training images that contain varying faces and gazes.

Since AAM is constructed statistically from training images, some elements of vector $\mathbf{c}$ represent the information related to the variance in face and gaze orientation. Therefore, this parameter vector $\mathbf{c}$ is employed as the feature parameter for the estimation of gaze orientation because parameter vector $\mathbf{c}$ is thought to be linearly associated with the displacement of the feature points caused by changes in head pose and gaze orientation.

In the AAM framework, shape vector $\mathbf{s}$ and texture vector $\mathbf{g}$ of the face are represented as shown in Eq. (1) and Eq. (2), respectively. In particular, shape vector $\mathbf{s}$ indicates the coordinates of the feature points, and texture vector $\mathbf{g}$ indicates the gray-level of the image within the shape,

$$\mathbf{s}(\mathbf{c}) = \bar{\mathbf{s}} + \mathbf{P_s}\mathbf{W_s}^{-1}\mathbf{Q_s}\mathbf{c} \tag{1}$$

$$\mathbf{g}(\mathbf{c}) = \bar{\mathbf{g}} + \mathbf{P_g}\mathbf{Q_g}\mathbf{c} \tag{2}$$

where $\bar{\mathbf{s}}$ and $\bar{\mathbf{g}}$ are the mean shape and mean texture of training images, respectively. $\mathbf{P_s}$ and $\mathbf{P_g}$ are a set of orthogonal bases of shape and texture variation, respectively. $\mathbf{Q_s}$ and $\mathbf{Q_g}$ are eigen matrices (including the eigenvectors). $\mathbf{W_s}$

is a diagonal weight matrix for each shape parameter, allowing for the difference in units between the shape and texture models. $\mathbf{c}$ is a vector of parameters controlling both the shape and gray-levels of the model.

In this paper, AAM is constructed using 43 shape points as shown in Fig. 2.

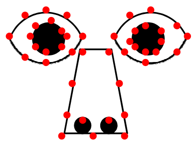

Fig. 2. 43 feature points for construction of AAM.

On the test image $\mathbf{I}$, the goal of the AAM search is to minimize the error $\mathbf{e}(\mathbf{p}, \mathbf{c})$ as shown in Eq. (3) with respect to parameter vector $\mathbf{c}$ and pose parameter vector $\mathbf{p}$.

$$\mathbf{e}(\mathbf{p}, \mathbf{c}) = || \ \mathbf{g}(\mathbf{c}) - \mathbf{I}(\mathbf{W}(\mathbf{p})) \ || \tag{3}$$

where $\mathbf{W}$ denotes the Affine warp function, $\mathbf{p}$ denotes the pose parameter vector for Affine warp (translation, scale, rotation), and $\mathbf{I}(\mathbf{W}(\mathbf{p}))$ indicates the Affine-transformed image controlled by the pose parameter $\mathbf{p}$ on the test image $\mathbf{I}$. $\mathbf{g}(\mathbf{c})$ is given in Eq. (2).

Thus, we can extract the most optimized parameter vector $\mathbf{c}$ as feature parameters from the test image.

2.3 Regression Analysis and Model Selection Method

The head pose and gaze orientation are estimated by regression analysis using the feature parameters extracted by AAM. In order to estimate horizontal face orientation ϕ, Cootes proposed a face rotation model [9]. In this paper, we propose a nobel method for simultaneous estimation of horizontal face orientation ϕ and horizontal relative gaze orientation θ based on the relationship between the displacement of feature points and rotation angles ϕ or θ by expanding Cootes's method.

In the proposed method, the regression formula to estimate horizontal facial orientation ϕ and horizontal gaze orientation θ can be represented by Eq. (4).

$$\mathbf{y} = \mathbf{a_0} + \mathbf{A}\mathbf{c} \tag{4}$$

where $\mathbf{y} = \left(\phi, \acute{\theta}\right)^T \in \mathbf{R}^{2\times1}$ is the vector of the objective variable. $\acute{\theta}$ is the total orientation of facial orientation ϕ and gaze orientation θ ($\acute{\theta} = \phi + \theta$), which means the horizontal gaze orientation relative to the image plane. $\mathbf{a_0} = \left(a_{1,0}, a_{2,0}\right)^T \in \mathbf{R}^{2\times1}$ is the constant vector of regression. $\mathbf{c} = \left(c_1 \ldots c_d\right)^T \in$

$\mathbf{R}^{d\times 1}$ is the parameter vector (explanatory variable) as given in Eq. (1) and Eq. (2). $\mathbf{A} \in \mathbf{R}^{2\times d}$ is the matrix of the regression coefficients as given in Eq. (5).

$$\mathbf{A} = \begin{pmatrix} a_{1,1} \dots a_{1,d} \\ a_{2,1} \dots a_{2,d} \end{pmatrix} \tag{5}$$

where d is the dimension of parameter vector $\mathbf{c}$.

Some components of parameter vector $\mathbf{c}$ are thought to be unessential when estimating horizontal facial orientation and horizontal gaze orientation because they sometimes cause over estimation when learning the regression coefficient matrix $\mathbf{A}$.

To solve this problem and improve the precision of this method, in this paper, the model selection method is employed to select the most suitable formula. In Eq. (4), for example, $\acute{\theta}$ can be represented as in Eq. (6).

$$\acute{\theta} = a_{2,0} + \sum_{i=1}^{d} a_{2,i}c_i + \epsilon(0, \sigma^2) \tag{6}$$

where estimation error ϵ is assumed to have Gaussian distribution with mean 0 and variance σ^2. We want to select only the essential components of parameter vector $\mathbf{c}$, but there are many combinations of the components. Therefore, we make S_k denote a set of the components among the following $2^d - 1$ sets.

$$\begin{aligned} S_1 &= \{c_1\} \\ S_2 &= \{c_2\} \\ S_3 &= \{c_1, c_2\} \\ &\vdots \\ S_{2^d-1} &= \{c_1, \dots, c_d\} \end{aligned}$$

Then, the regression formula for S_k given in Eq. (6) can be represented as follows.

$$\acute{\theta}_k = a_{2,0} + \sum_{i\in S_k} a_{2,i}c_i + \epsilon(0, \sigma_k^2) \tag{7}$$

After learning the regression coefficients using the least squared method among $k = 2^d - 1$, the least scored model is selected as the most suitable model.

Akaike Information Criterion[10] (AIC) is one of model selection methods in regression analysis, which indicates generalization capability of regression formular using training data.

The maximum log-scaled likelihood $l(\Theta_k; X)$ and the degrees of freedom of the model are evaluated by AIC as shown in Eq. (8).

$$\mathrm{AIC}_k = -2l(\Theta_k; X) + 2\dim(\Theta_k) \tag{8}$$

$$l(\Theta_k; X) = -\frac{n}{2}(1 + \log(2\pi\sigma_k^2)) \tag{9}$$

where n denotes the number of training images. Θ_k denotes the model parameters given in Eq. (7), and maximum log-scaled likelihood $l(\Theta_k; X)$ is assumed to be given in Eq. (9). The lower the AIC score, the better the evaluation of the model. This means that AIC gives an answer with a trade-off between the complexity of the model and the variance σ_k^2 of the fitting error ϵ to the training image set X as given in Eq. (9).

In a similar way, Minimum Description Length[11] (MDL), and Bayesian Information Criterion[12] (BIC) are respectively defined as shown in Eq. 10, and Eq. 11.

$$\mathrm{MDL}_k = -l(\Theta_k; X) + \frac{\dim(\Theta_k)\ln n}{2} \tag{10}$$

$$\mathrm{BIC}_k = -2l(\Theta_k; X) + \dim(\Theta_k)\ln n \tag{11}$$

Thus, matrix $\mathbf{A}$ and vector $\mathbf{a_0}$ are trained in the above mentioned methods. Horizontal gaze orientation relative to the image plane can be estimated as shown in Eq. (12) using the parameter vector $\mathbf{c}$ of the test image.

$$k = \arg\min_k \begin{cases} \mathrm{AIC}_k \\ \mathrm{MDL}_k \\ \mathrm{BIC}_k \end{cases} \tag{12}$$

3 Experiment

To confirm the validity of our method in estimating horizontal gaze orientation, we conducted the following experiment.

3.1 Experimental Conditions

Since there was no open dataset with variation of face and gaze orientation, we prepared a dataset by asking each subject to look at each of the markers on a wall in turn. The markers were placed horizontally on the wall at every 5 degrees. The variations of head pose and gaze orientation were in the horizontal direction and ranged from approximately -20 degrees to +20 degrees relative to the front. The dataset contained 4 subjects, 63 training images and 252 test images with 640 x 480 pixels for each subject, as shown in Table 1.

In this experiment, we used AAM parameters with up to 95% cumulative contribution ratio. In fact, the number of dimensions were about 10-20.

AAM construction and regression analysis were performed for each subject. The proposed method was evaluated by comparing it with the method proposed by Ishikawa et al (a conventional method). The horizontal gaze orientation estimation method was evaluated by means of absolute error degree (MAE).

Moreover, we conducted this experiment with the purpose of showing the validity and the contribution of the two unique points of the proposed method.

Table 1. Overview of our dataset for each subject ("Training" means the number of training images, and "Test" means the number of test images.)

Face [deg]	Gaze [deg]	Training	Test
0(Frontal)	± 20, ± 15, ± 10, ± 5, 0	9	36
5	+20, +15, ± 10, ± 5, 0	7	28
-5	-20, -15, ± 10, ± 5, 0	7	28
+10	+20, +15, ± 10, ± 5, 0	7	28
-10	-20, -15, ± 10, ± 5, 0	7	28
+15	-20, -15, ± 10, ± 5, 0	7	28
-15	-20, -15, ± 10, ± 5, 0	7	28
+20	-20, -15, -10, ± 5, 0	6	24
-20	-20, -15, -10, ± 5, 0	6	24
In total		63	252

At first, we compared the proposed method with the Ishikawa et al.'s method. In our method, we didn't use any model selection methods. On the other hand, we gave true coordinates of AAM shape points in Ishikawa's method because it required a lot of time for us to implement 3D AAM. Through this comparison, the validity of our method can be confirmed.

Next, we compared the difference between simultaneous method and sequential method. In s̈equential;, the angle θ of the gaze in relation to the face is described in Eq. (4). After regression analysis, total horizontal gaze orientation $\acute{\theta}$ is computed as $\phi + \theta$. Through this comparison, the validity of "simultaneous" estimation can be confirmed.

Finally, we compared AIC with other methods of model selection. MDL (Minimum Description Length) [11] and BIC (Bayesian Information Criterion) [12] are well-known methods for selecting the model in recent years. Thus, the validity of "model selection by AIC" can be confirmed by comparing it with MDL and BIC.

3.2 Results

Fig. 3 shows the experimental results. We can confirm the validity of s̈imultaneous m̈ethod. The graph shows the average estimation error [deg]. Though the difference among these methods seems small, a significant difference is confirmed with significance level of 95%.

This graph shows that our approach contributes the improvement of the gaze estimation error from 4.2 [deg] to 2.7 [deg]. I think there is no critical difference among three model selection methods. But it is important to reduce the dimension of the feature vector using model selection method.

Next, We analyzed the performance of our approach (e.g. AIC) in each face direction. Fig. 4, Fig. 5, Fig. 6, and Fig. 7 respectively show that the mean estimation error in each face angle and each subject. From these graph, we can confirm the face angle robustness of the proposed method.

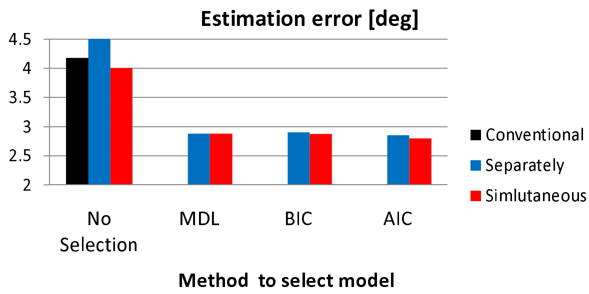

Fig. 3. Experimental results (Mean estimation error).

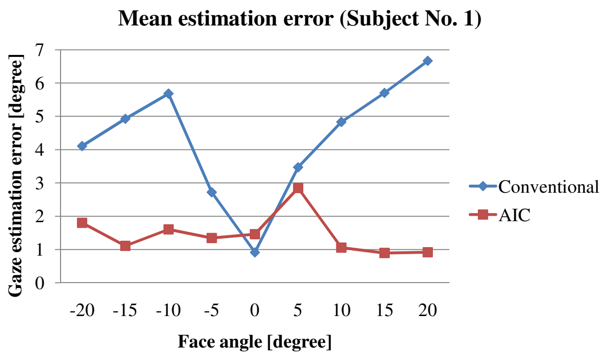

Fig. 4. Experimental results (estimation error) of subject No. 1 in each face angle.

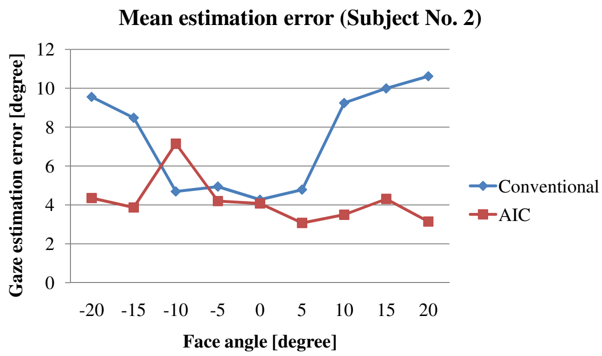

Fig. 5. Experimental results (estimation error) of subject No. 2 in each face angle

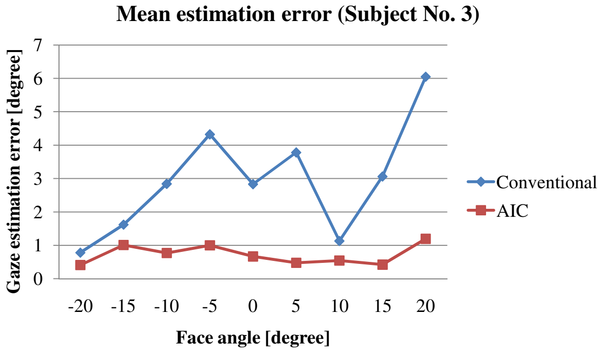

Fig. 6. Experimental results (estimation error) of subject No. 3 in each face angle

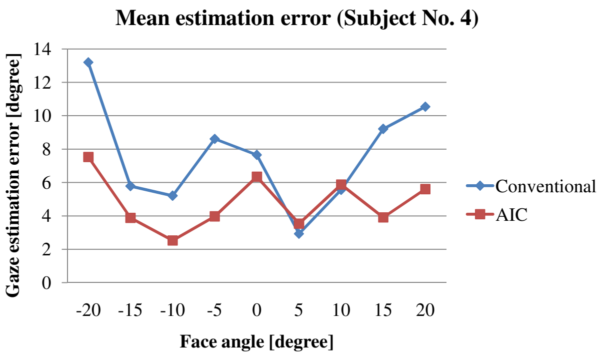

Fig. 7. Experimental results (estimation error) of subject No. 4 in each face angle

From these experimental results, the validity of the proposed method to estimate horizontal gaze orientation from monocular images was confirmed. Also, it was confirmed that the model selection by AIC contributes the most to reducing the degree of error.

4 Conclusion

In this paper, the nobel method was proposed in which horizontal gaze orientation is estimated by using model selection method to select the necessary parameters from to the AAM parameters and then carrying out regression analysis on those paramters. This method contributes to the improvement of horizontal gaze estimation error from 4.2 [deg] to 2.7 [deg].

In near future research, we will address the problem of AAM adaptation to an unseen subject for a wide range of gaze estimation applications.

References

1. M. Miyahara, M. Aoki, T. Takiguchi and Y. Ariki, "Tagging Video Contents with Positive/Negative Interest, " The 14th International Multimedia Modeling Conference, pp. 210-219, 2008
2. D. Pang, A. Kimura, T. Takeuchi, and J. Yamato , "A STOCHASTIC MODEL OF SELECTIVE VISUAL ATTENTION WITH A DYNAMIC BAYESIAN NETWORK, " IEEE International Conference on Multimedia and Expo, pp. 1073-1076, 2008
3. T. Ohno, N. Mukawa, and A. Yoshikawa, "FreeGaze: gaze tracking systems for everyday gaze interaction, " Proceedings of the symposium on Eye tracking research & applications, pp. 125-132, 2002
4. B. Lucas, and T. Kanade, "An interactive image registration technique with an application to stereo vision, " Proc Int'l Joint Conference on Atrificial Intelligence, pp. 674-679, 2005
5. H. Yamazoe, A. Utsumi, T. Yonezawa, and S. Abe, "Remote Gaze Estimation with a Single Camera Based on Facial-Feature Tracking without Special Calibration Actions, " Proceedings of the symposium on Eye Tracking Research & Applications Symposium, pp. 245-250, 2008
6. T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive Driver Gaze tracking with Active Appearance Models, " In Proc. 11th World Congress in Intelligent Transport Systems, 2004
7. P. Viola, and M. J. Jones, "Robust Real-Time Face Detection, " International Journal of Computer Vision, Vol.2, pp. 137-154, 2004
8. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models, " European Conference on Computer Vision, pp. 484-498, 1998
9. T. F. Cootes, K. Walker, and C. J. Taylor, "view-based Acitve Appearance Models, " Forth IEEE Conference on Automatic Face and Gesture Recognition, pp. 227-232, 2000
10. A. Hirotugu, "A new look at the statistical model identification, " IEEE Transactions on Automatic Control, Vol.19, pp. 716-723, 1974
11. J. Rissanen, "Infomation and Complexity in Statistical Modeling, " Springer, 2007
12. A. D. R. McQuarrie, and C. L. Tsai "Regression and Time Series Model Selection, " World Scientific, 1998