2012 International Workshop on Nonlinear Circuits,
Communications and Signal Processing
NCSP'12, Honolulu, Hawaii, USA. March 4-6, 2012

Gaze Estimation Using 3D Active Appearance Models

Yukari Nakamatsu, Tetsuya Takiguchi and Yasuo Ariki

†Graduate School of Engineering, Kobe University
1-1 Rokkodai, Nada-ku, Kobe, 657-8501 Japan
Phone/FAX:+81-78-803-6570
Email: nakamatsu@me.cs.scitec.kobe-u.ac.jp Email: takigu@kobe-u.ac.jp Email: ariki@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 gaze orientation from images obtained from inexpensive cameras (such as web cameras) based on 3D active appearance models (3D-AAM), where the 3D-AAM is used to extract the coordinates of the feature points and the gaze orientation. The proposed 3D-AAM method is able to estimate gaze orientation using only two training 3D images. The experimental results show that the proposed method is able to improve gaze estimation accuracy and decrease the number of training images required, in comparison to the conventional method.

1. Introduction

Visual information makes up a large percentage of human perception. Therefore, gaze estimation is expected to be applied to human interaction and human interest estimation. Gaze estimation research associated with human safety in driving has also been carried out [1]. Recently, a technique for estimating gaze has also been used for digital signage, in which the technology is used to determine who watches a screen or showcase, and which portion of the screen they are most interested in.

Many kinds of methods for estimating gaze have been proposed. One approach uses a special device (such as an infrared camera). This approach can estimate gaze with a high degree of accuracy, but it is necessary to install the device on one's head, and it is very expensive. Other approaches are estimating gaze by using image processing, for example, where a 3D-eyeball model is applied to eye images [2]. In those approaches based on image processing, it is easy to prepare cameras and not to put a strain on the user. However, in many approaches, the gaze estimation precision is not very high because many training data are needed. In this paper, the proposed method can estimate the gaze orientation using only two images based on 3D-AAM, where the number of training data is less than that required by conventional methods.

2. Outline of the Proposed System

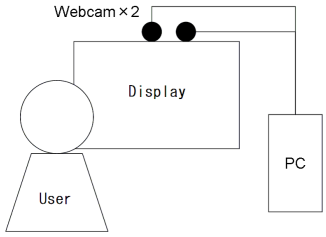

Figure 1: Experimental environment

Fig. 1 shows the experiment environment used in this paper. A subject looks at each of the marks on a computer screen, where there is no assumption for the face direction of the subject. The 3D position of the subject's face is calculated using two webcams on the monitor in order to train the 3D-AAM. Then, only one webcam is used to estimate the gaze orientation in the test process.

The flow of image processing is shown in Fig. 2. In the proposed method, the facial area for the training and test image is detected using AdaBoost based on Haar-like features. Next, the feature parameters of the face and eyes are extracted using the 3D-face-AAM on the detected facial area and the 3D-eye-AAM on the detected eye area, respectively. Finally, the 3D-eye-AAM fitting is calculated in order to calculate the gaze orientation.

3. Active Appearance Model

Cootes proposed an AAM to represent shape and texture variations of an object with a low dimensional parameter vector $\mathbf{c}$ [3]. Vector $\mathbf{c}$ can represent various facial images with arbitrary face and gaze orientation using training images that contain varying faces and gazes. Since an AAM is constructed statistically from training images, some elements of vector $\mathbf{c}$ represent the information related to the variance in

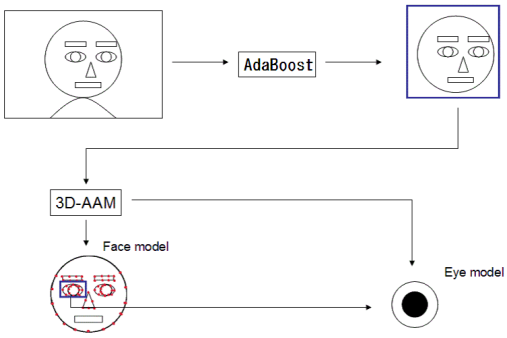

Figure 2: Flow of image processing

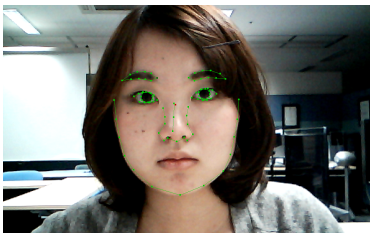

Figure 3: Feature points of 2D-AAM

face and gaze orientation [4]. Therefore, this parameter vector **c** is employed as the feature parameter for estimating the gaze orientation because parameter vector **c** is considered to be linearly associated with the displacement of the feature points caused by changes in the head pose and gaze orientation.

In the AAM framework, shape vector **s** and texture vector **g** of the face are given as follows:

$$\mathbf{s} = (x_1, y_1, x_2, y_2, ..., x_n, y_n)^T, \quad \mathbf{g} = (g_1, g_2, ..., g_m)^T \quad (1)$$

where the shape vector **s** indicates the coordinates of the feature points, and the texture vector **g** indicates the gray-level of the image within the shape. In this paper, the AAM is constructed using 89 shape points as shown in Fig. 3.

Next, principal component analysis (PCA) is applied to the training data, and the normal orthogonal matrices, $\mathbf{P_s}$ and $\mathbf{P_g}$, are obtained. Using the obtained matrices, the shape vector and the texture vector can be approximated as follows:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{P_s b_s} \quad (2)$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P_g b_g} \quad (3)$$

where $\bar{\mathbf{s}}$ and $\bar{\mathbf{g}}$ are the mean shape and mean texture of the training images, respectively. $\mathbf{b_s}$ and $\mathbf{b_g}$ are the parameter of variation from the average. Further PCA is applied to the vector **b** as follows:

$$\mathbf{b} = \begin{pmatrix} \mathbf{W_s b_s} \\ \mathbf{b_g} \end{pmatrix} = \begin{pmatrix} \mathbf{W_s} {\mathbf{P_s}}^T (\mathbf{s} - \bar{\mathbf{s}}) \\ {\mathbf{P_g}}^T (\mathbf{g} - \bar{\mathbf{g}}) \end{pmatrix} = \begin{pmatrix} \mathbf{Q_s} \\ \mathbf{Q_g} \end{pmatrix} \mathbf{c} \quad (4)$$

where $\mathbf{W_s}$ is a diagonal weight matrix for each shape parameter, allowing for the difference in units between the shape and texture models. $\mathbf{Q_s}$ and $\mathbf{Q_g}$ are the eigen matrices (including the eigenvectors). **c** is a vector of parameters controlling both the shape and gray-levels of the model. Finally, the shape and texture are approximated as functions of **c**.

$$\mathbf{s}(\mathbf{c}) = \bar{\mathbf{s}} + \mathbf{P_s W_s^{-1} Q_s c} \quad (5)$$

$$\mathbf{g}(\mathbf{c}) = \bar{\mathbf{g}} + \mathbf{P_g Q_g c} \quad (6)$$

Using parameter **c**, it is possible to control variations in shape and texture of the AAM, but it is not possible to express the position of the face in the image, the size of the face or the head pose. The pose parameter p is defined as the global posture change as follows:

$$p = [roll \;\; scale \;\; trans\_x \;\; trans\_y] \quad (7)$$

where $roll$ indicates the rotation to the model plane, $scale$ indicates the size of the model, and $trans\_x$ and $trans\_y$ indicate the translation among x and y, respectively.

The goal of the AAM search is to minimize the error $e(\mathbf{p}, \mathbf{c})$ on the test image I as shown in Eq. (6) with respect to **c** and **p**,

$$e(\mathbf{p}, \mathbf{c}) = || \; \mathbf{g}(\mathbf{c}) - I(F(\mathbf{p})) \; || \quad (8)$$

where F denotes the affine warp function, and $I(F(\mathbf{p}))$ indicates the affine transformed image controlled by the pose parameter **p** on the test image I. $\mathbf{g}(\mathbf{c})$ is given in Eq. (6). Thus, we can extract the most optimized **c** from the test image.

4. 3D Active Appearance Model

The 3D-AAM [1, 5] is described in this section. In this paper, the 3D-AAM is used to extract the face feature and estimate the gaze orientation. The AAM including various fluctuation components (images) can be constructed by using the facial images, including various changes in the training data (such as a left-side face image or a right-side one, and an upturned or a downturned face, etc). Moreover, the more training data there is, the higher the accuracy of the model.

The method for estimating the direction of the face and gaze has been proposed in [4], where Gaussian process regression is introduced in order to deal with the relation between gaze direction and the parameters obtained using 2D-AAM, and many training images of various directions of the face and gaze are required. However, it is no simple matter to prepare a large number of facial images for each person before the system is used.

An arbitrary face direction image is created by using a face shape of three dimensions and a face texture in the front because the variation of the head pose can be expressed by a

geometrical transformation of the 3D-AAM. Therefore, the 3D-AAM does not require much training data.

The shape parameter is expanded into 3D using z from stereo matching as shown in Eq. (9).

$$\mathbf{s} = (x_1, y_1, z_1, x_2, y_2, z_2, ..., x_n, y_n, z_n)^T \tag{9}$$

The 2D pose parameter in Eq. (7) is expanded into 3D adding yaw and $pitch$ as shown Eq. (10).

$$p = [yaw \ \ pitch \ \ roll \ \ scale \ \ trans\_x \ \ trans\_y] \tag{10}$$

The moving variations of these parameters are shown in Fig. 4.

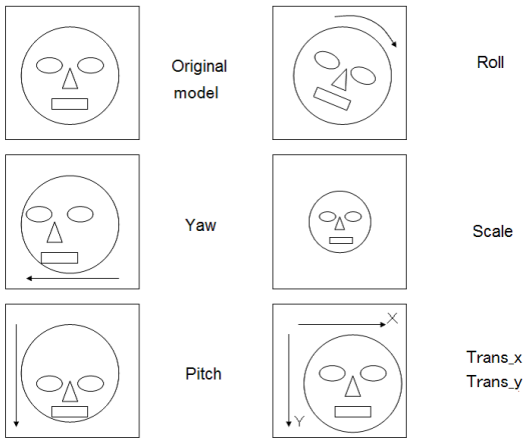

Figure 4: 3D pose parameter

Using the six parameters, the 2D-AAM can be expanded into the three dimensions, and it is able to transform the model to all directions, angles, and positions. The transformation of the shape using this pose parameter is given as follows:

$$p_a = Trans \cdot Scale \cdot RotZ \cdot RotY \cdot RotX \cdot p_b \tag{11}$$

where p_b indicates the shape coordinate before transformation. Each transformation matrix is given by Eq. (12) $\sim$ (16).

$$Trans = \begin{pmatrix} 1 & 0 & 0 & trans\_x \\ 0 & 1 & 0 & trans\_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{12}$$

$$Scale = \begin{pmatrix} scale & 0 & 0 & 0 \\ 0 & scale & 0 & 0 \\ 0 & 0 & scale & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{13}$$

$$RotX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha * \pi/180) & -\sin(\alpha * \pi/180) & 0 \\ 0 & \sin(\alpha * \pi/180) & \cos(\alpha * \pi/180) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{14}$$

$$RotY = \begin{pmatrix} \cos(\beta * \pi/180) & 0 & \sin(\beta * \pi/180) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\beta * \pi/180) & 0 & \cos(\beta * \pi/180) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{15}$$

$$RotZ = \begin{pmatrix} \cos(\gamma * \pi/180) & -\sin(\gamma * \pi/180) & 0 & 0 \\ \sin(\gamma * \pi/180) & \cos(\gamma * \pi/180) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16}$$

As the shape in 3D-AAM consists of three dimensions and the input image consists of two dimensions, it is necessary to project the 3D space into the 2D space when calculating the error between the input model and the 3D model. Using the projection function G from the 3D space into the 2D space, Eq. (8) is calculated as follows. Then, the optimizing parameter is calculated using the same approach used for 2D-AAM.

$$e(\mathbf{p}, \mathbf{c}) = || \ \mathbf{g}(\mathbf{c}) - I(G(F(\mathbf{p}))) \ || \tag{17}$$

5. Gaze Estimation

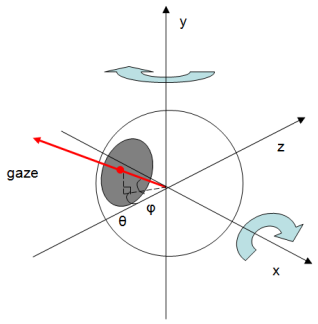

Figure 5: Eye-model

In order to estimate the gaze, an eye model based on the 3D-AAM is trained using the image data of the eye region obtained from the 3D-AAM associated with the face, where the 3D-AAMs of the face and the eye are constructed using 89 shape points as shown in Fig. 3 and 16 shape points, respectively. The eye model used in the proposed method is shown in Fig. 5. The eyeball is treated as a shape of an ideal ball, and the pupil is considered to be a circle on the eyeball as shown in Fig. 5. The center of the pupil is assumed to go through the center of the eye, and corresponding to an optical axis of the eyeball. It is assumed that the optical axis corresponds to the subject's gaze. In Fig. 5, θ indicates the rotation in the horizontal direction, and ϕ indicates the rotation in the vertical direction. The 3D AMM of the eye is able to transform the model to all directions, angles, and positions. The

transformation of the shape is given as follows using a similar approach to Eq. (11):

$$p_d = Trans \cdot Scale \cdot RotZ' \cdot RotY' \cdot RotX' \cdot p_c \quad (18)$$

where p_c indicates the shape coordinate before transformation. Each transformation matrix is given by Eq. (12) $\sim$ (16) when substituting θ and ϕ for yaw and $pitch$, respectively. We used the $scale$ and the diameter of the eyeball estimated using the 3D-AMM of the face.

6. Experiments

6.1. Experiment Conditions

We recorded the gaze image data in the environment shown in Fig. 1. One subject looks at a mark on a computer screen. Two webcams on the monitor captured two images (with 640 $\times$ 480 pixels) for training the 3D-AAMs of the face and eye. The distance between the center of their lenses was 100 mm.

The screen size was 375 $\times$ 300 mm. The marks were placed on the screen in 5-degree increments from the center mark (directly in front of the subject), and the variations of the gaze orientation were ranged horizontally from -15 degrees to +15 degrees and vertically from -10 degrees to +10 degrees. The total number of test data was 35. Only the left-side camera was used for testing.

6.2. Experiment Results

The average error of the horizontal direction and the vertical direction are shown in Fig. 6 and Fig. 7, respectively. As shown in the two figures, the proposed method based on the 3D-AAM results in a similar error degree when compared with the 2D-AAM [4] which are trained using 315 image data captured with one webcam. The experiment results also show that the 3D-AAM decreases the degree of error when compared with 2D-AAMs that are trained using 81 image data and 2 image data, respectively. As the proposed method uses only two image data captured with two webcams, it can decrease the number of training images compared to the conventional method.

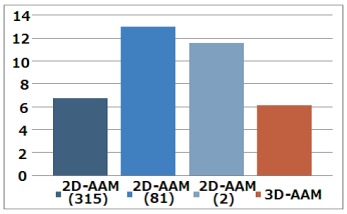

Figure 6: Horizontal error [degree]

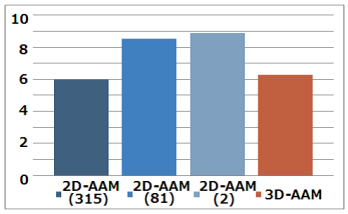

Figure 7: Vertical error [degree]

7. Conclusion

In this paper, we have presented a gaze estimation method, where the 3D-face AAM and the 3D-eye-AAM are used. It was found that the proposed method can estimate the gaze orientation that is independent of the direction of the face. As a result of the experiment, we found our approach is effective in decreasing the number of training data required in comparison with the conventional method. In the near future, we will carry out research aimed at addressing the problems associated with using an AAM on the small region of the eye and the AAM adaptation to an unknown subject for a wide range of gaze estimation applications.

Acknowledgment: This research was supported in part by MIC SCOPE.

References

[1] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, “Passive Driver Gaze Tracking with Active Appearance Models,” Proceedings of the 11th World Congress on Intelligent Transportation Systems, 2004

[2] 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

[3] T.F. Cootes, G.J. Edwards, and C.J. Taylor, “Active appearance models,” Proceedings of ECCV, pp. 484-498, 1998

[4] M. Takatani, Y. Ariki, and T. Takiguchi, “Gaze Estimation Using Regression Analysis and AAMs Parameters Selected Based on Information Criterion,” International Workshop on Gaze Sensing and Interactions in conjunction with ACCV2010, pp. 1-10, 2010

[5] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade, “Real-Time Combined 2D+3D Active Appearance Models,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535-542, 2004