An Adaptive Active Shape Model for Eye Shape Detection

Mohammad ALDIBAJA¹ and Shinichi SUZUKI¹

¹ Department of Mechanical Engineering, Toyohashi University of Technology, Toyohashi 441-8580, Japan

(Received 11 January 2014; received in revised form 9 May 2014; accepted 7 June 2014)

Abstract
This paper proposes an improved Active Shape Model (ASM) to detect the eye shape. The initial estimation of the eye shape and the poor description of its pixel value variation are considered as common problems in ASM. We employ Principal Component Analysis (PCA) to suggest an initial estimation of the eye shape based on understanding the eye structure in an eye image. In addition, the properties of PCA in pixel values-matching and data reduction are used to describe the pixel value variation around the eye shape and to search on the proper locations of the eye shape in the eye image. The experimental results verified the effectiveness of the proposed method compared to the standard ASM.

Key words
Active Shape Model, Principal Component Analysis, Shape Detection, Eye Features

1. Introduction
Eye features extraction and analysis are mainly considered as a key point for implementing a noncontact human-computer interface system [1]. The applications can be found in recognition systems of human face [2]. It can primary be used to estimate the locations of the other facial features [3]. For example, the face rotation angle is easily determined using the two eye corners. Furthermore, by measuring the distance between pupil and some certain points on the eye shape, the gaze direction can be precisely determined [4]. Many methods in the field of object shape detection have been proposed [5-8]. Among these methods, Active Shape Model (ASM), proposed by Cootes and et al. [6], is the simplest and most effective strategy can be used for detecting the eye shape. The eye shape is described by a set of points that represent the upper and lower eyelids and two eye corners. ASM is an iterative technique that analyzes and models the point distribution (PDM) of a set of training eye shapes. The modeling operation uses a statistical method called Principal Component Analysis (PCA) [9]. Moreover, ASM models the variation of the gray texture (pixel values) at each eye shape point using a set of training eye images. Based on the shape and gray texture models, ASM searches on the correct location of the eye shape in a given image.

Various approaches have been suggested to improve the performance of ASM [10-13]. For example, Cootes used a local gray-level pyramid in order to increase the robustness of the searching operation against illumination changes [10]. Sukno and et al. [11] used a non-linear classifier to label the shape points near the object boundary. The advantage is that this classifier makes the description of pixel values around the shape points invariant to rigid transformation.

However, most of these methods follow the standard ASM strategy in suggesting a global shape called mean shape as an initial estimation of the object shape. Mean shape restricts ASM to process the spatial changes of the object. This may lead to increase the deformations of the detected shape. Moreover, these methods use a 1-D profile of pixel values to model the gray texture around the shape points. A 1-D profile has low capability to detect the correct position of the object shape because it possesses a few pixels in one dimension.

This paper proposes an approach to overcome the above mentioned problems. First, using a set of training eye images and corresponding eye shapes, we utilize object recognition technique based-PCA to suggest an initial estimation of the exact eye shape in a given image. This procedure gives ASM more flexibility to deal with wide range of change in eye profile parameters such as rotation, scale, transformation and eye state (open, nearly closed and in-between). Second, we extend the 1-D profile to 2-D profile called sub-image. This procedure enriches the description of pixel values at eye shape points and increases the robustness of searching operation.

The experimental results verified that the proposed method outperforms the standard ASM technique in increasing the accuracy level and reducing time consumption.

2. Standard Active Shape Model
ASM is divided into two stages; training and detection. In training stage, ASM is charged with a set of eye images and corresponding eye shapes (database). Point distribution and gray texture of these eye shapes are modeled. In detection stage, an eye image, testing image, is presented and ASM tries to detect the eye shape based on the gray texture and point distribution models. More details on ASM stages are given in the following sections.

2.1 Training stage
2.1.1 Eye shape modeling
Figure 1(a) shows a sample of the eye images in database. The exact eye shape is labeled manually as highlighted by N white points. The point distribution represents the locations of two eye corners and eyelids. The corresponding x-y coordinates are contained in a vector X.

\[ X = \left( x_1, x_2, \ldots, x_j, \ldots, x_n, y_1, y_2, \ldots, y_j, \ldots, y_n \right)^T \]  

where \( x_j \) and \( y_j \) are the x-y coordinates of \( j^{th} \) point; respectively and \( T \) implies the vector transpose operation.
Fig. 1 ASM training stage: (a) An eye image in database and manually labeled eye shape points. (b) Gray texture around shape points. (c) 1-D profiles of the first point. (d) Mean profile of the first point

Suppose that the number of eye images in database is $M$. This means that the output of labeling operation is $M$ eye shape vectors. In order to model the statistical variation of point distribution among these vectors, Euclidean similarity must be minimized using an alignment algorithm such as Procrustes analysis [14]. The aligned eye shape vectors are then arranged in a matrix $\Phi$.

$$
\Phi = [\hat{X}_1, \hat{X}_2, ..., \hat{X}_i, ..., \hat{X}_M]^T
$$

where $\hat{X}_i$ is the aligned eye shape vector of $i^{th}$ image. The mean eye shape vector $\bar{X}$ and the covariance matrix $C$ are then calculated as follows:

$$
\bar{X} = \frac{1}{M} \sum_{i=1}^{M} \hat{X}_i \quad \text{and} \quad C = \frac{1}{M} \Phi \Phi^T
$$

The $M$ eigenvalues and $M$ eigenvectors of the covariance matrix $C$ are computed in order to obtain the eye shape model equation.

$$
X_i = \bar{X} + \hat{X}_i = \bar{X} + \Omega B_i
$$

where $B_i=(b_1,b_2, ..., b_M)$ is a vector, called decomposition vector, represents the contribution of each eigenvector in constructing $X_i$.

2.1.2 Gray texture modeling around eye shape points

Figure 1(b) highlights 1-D profiles by white lines at the labeled eye shape points. A line at $i^{th}$ point is in the normal direction and perpendicular to the line passing between the former and next points. The $K$ pixels covered by a white line, as enlarged in a white rectangle, are contained in a vector $g$.

$$
g_{ij} = (I_1, I_2, ..., I_{i-1}, I_i, ..., I_K)
$$

where $g_{ij}$ is a 1-D profile of $i^{th}$ point in $j^{th}$ eye image in database and $I_i$ is the gray value of $i^{th}$ pixel.

Suppose that the database contains $M$ eye images. This means that an $i^{th}$ eye shape point has $M$ 1-D profiles as shown in Fig.1(c-d). In order to describe the variation of pixel values exhibited by these profiles, a mean profile vector $\bar{g}_i$ and a covariance matrix $G_i$ are calculated in the same manner of Eq.(3). Thus, this operation is repeated on each eye shape point and we end up with $N$ covariance matrices and corresponding $N$ mean profiles.

2.2 Detection stage

The flowchart of ASM detection strategy is shown in Fig.2(a). An eye image is presented to detect the exact eye shape. Figure 2(b) shows the exact shape depicted by a white dashed curve. ASM uses the mean eye shape $\bar{X}$ as an initial estimate of the exact shape. Mean eye shape is fitted to the given image by a black curve. A searching operation is then applied to attract the mean eye shape points to the exact shape. Figure 2(c) artificially demonstrates the searching strategy at $2^{nd}$ point of the mean eye shape. A searching line is created in the normal direction of $1^{st}$ and $3^{rd}$ points. ASM searches along this line on a segment $\bar{X}$ whose pixel values minimizes Eq.(6) using $2^{nd}$ mean profile $\bar{g}_2$ and $2^{nd}$ covariance matrix $G_2$ that calculated in the training stage. The location $u$ of this segment is expected to be at the target point (exact eye shape) or close to it.

$$
\mathcal{F}_2(g_u) = (g_u - \bar{g}_2)^T G_2^{-1} (g_u - \bar{g}_2)
$$

After applying the searching procedure on $N$ points of the mean eye shape, a new shape is obtained. This shape is expected to have many deformations. In order to reform
Fig. 3 Consequences of bad initial estimations: (a) Testing images. (b) Mean eye shape fitting. (c) Detected shapes by standard ASM

3. Adaptive Active Shape Model

ASM is able to detect an object shape in binary image domain [15]. On the other hand, many issues must be considered when the image is captured in gray domain, such as lighting conditions, shadows, rotation, scale, etc. In the following two sub-sections, we use PCA to improve both the initial eye shape estimation and the searching operation.

3.1 Adaptive initial estimation

3.1.1 Problem and proposed solution

The initial estimation of the exact eye shape plays a significant role in determining the behavior of searching operation as well as the number of iterations. ASM always uses mean eye shape as an initial estimate as shown in Fig. 2(a-b). Mean shape has fixed parameters such as scale, rotation and eye state (open) that makes ASM incapable to deal with changing of the eye profile.

Figure 3 demonstrates some consequences of bad initial estimations. Four eye images with different eye profiles are presented, Fig. 3(a). These images are used to show the performance of standard and improved ASM. The mean eye shape represented by white points is fitted over the images, Fig. 3(b). Figure 3(c) shows the corresponding final eye shapes detected by the standard ASM.

In the first image, a translation in $y$ coordinates occurs. The searching operation is not able to discriminate the correct locations of lower and upper eyelids because of the close distance. The second image shows the left upper points of the mean shape are very close to the eyebrow. Therefore, these points are attracted to the eyebrow instead of the upper eyelid in the final detection. In the third image, a large difference in rotation occurs and obviously most of searching lines do not house the exact eye shape initially.

3.1.2 Object recognition based-PCA

The comparison between images is the main problem in object recognition. As the database contains eye images only, we expect that they form a narrow cluster in the image domain as illustrated in Fig. 5. An eye image is represented by a point. This point moves within the cluster according to the change of a few eye parameters such as eye state, scale, rotation, pupil location. Based on this fact, the major and minor axes of the cluster describe different patterns of the eye parameters. Consequently, some of these patterns can be utilized to predict the state of the eye in a given image. PCA is utilized to describe the patterns by fitting the cluster by a set of eigenvectors that are in the directions of the maximum variation [9] as shown in Fig. 5.

The same strategy described in sub-section 2.1.1 is used for calculating mean vector, covariance matrix and corresponding eigenvectors with taking into account that we deal with pixel values. Therefore, each eye image in database is converted into a vector by concatenating each...
Fig. 5 Image domain, eye cluster and eigenvector space

\[ \Gamma_i = (I_{00}, I_{01}, \ldots, I_{jk}, \ldots, I_{m0}, I_{m1}, \ldots, I_{mn})^T \]  

(7)

where \( \Gamma_i \) is the vector of the \( i^{th} \) image in database, \( I_{jk} \) is the pixel value at \( j^{th} \) row and \( k^{th} \) column.

The \( M \) obtained eigenvectors are arranged in a matrix \( \Psi \) according to corresponding eigenvalues in descending order as partially shown in Fig. 6(a). The \( i^{th} \) eye image in database can be perfectly reconstructed using the mean eye image \( \bar{\Gamma} \) and a linear combination of eigenvectors as formed in Eq. (8). The decomposition vector \( D \) contains the projections of the eye image on each eigenvector as shown in Fig. 5.

\[ \Gamma_i = \bar{\Gamma} + \Psi D_i \]  

(8)

3.1.3 Understanding eye topological structure

PCA is employed to understand the topological structure of an eye rather than to recognize a specific person. The eye images in database are carefully chosen to accomplish as various scale and rotation of the eye as possible. These two parameters are expected to dominate the first few eigenvectors. Figure 6(a) proves this fact by showing in each row, the first, intermediate and last four eigenvectors; respectively. The first eigenvectors with large eigenvalues contain information relative to the eye structure. These eigenvectors are useful in understanding the eye structure. Eigenvectors with intermediate eigenvalues illustrate information that common to particular eye components such as pupil location and eye corners. Therefore, they can be used to give initial positions of these components. The last eigenvectors have small eigenvalues and represent the noise patterns in database and can be used to smooth the eye images. Obviously, the most significant information to understand the eye topological structure is encoded in the first eigenvectors. These eigenvectors are contained in a new matrix \( \Psi' \).

Eigenvectors with intermediate eigenvalues illustrate information that common to particular eye components such as pupil location and eye corners. Therefore, they can be used to give initial positions of these components. The last eigenvectors have small eigenvalues and represent the noise patterns in database and can be used to smooth the eye images. Obviously, the most significant information to understand the eye topological structure is encoded in the first eigenvectors. These eigenvectors are contained in a new matrix \( \Psi' \).

Figure 6(b), first row, shows the images in Fig. 3(a) reconstructed using \( \Psi' \). The most salient feature can be observed is the eye geometrical structure.

In the improved ASM, training stage, the \( M \) eye images in database are projected into the condensed eigenvector space \( \Psi' \) using Eq. (9). The corresponding decomposition vectors \( D \) are stored in a matrix called eye profile database.

\[ \hat{D}_i = \Psi'^{-1}(\Gamma_i - \bar{\Gamma}) \]  

(9)

In detection stage, when an eye image \( I_{\text{new}} \) is given, as the images in Fig. 3(a), it is projected into the eigenvector space \( \Psi' \) and represented by a point. The decomposition vector \( D_{\text{new}} \) of this point is obtained by Eq. (9). The decomposition vector \( D_{\text{new}} \) is then compared with those in the eye profile database respectively. The comparison is measuring the distance in the eigenvector space between the point represents \( I_{\text{new}} \) and those represent the eye images in database. Based on the comparison result, an image is taken from database as shown in Fig. 6(b), second row. The geometrical structures of the taken images are expected to be similar to those in Fig. 3(a). Therefore, the manually labeled eye shapes of the taken images are used as initial estimations as shown in Fig. 6(b), last row.

3.2 Adaptive searching operation

Searching operation is used iteratively to detect the exact eye shape as flowcharted in Fig. 2(a). It must be robust to prop the decision of the initial estimation in the first iteration and to efficiently reduce the eye shape deformations in the next iterations. Thus, searching operation is designed to detect the exact eye shape as less number of iterations as possible. Figure 1(b) shows a 1-D profile lies in one direction. Obviously, it does not provide significant information on the gray texture variation at the shape points. The few pixels contained are incapable to deal with various lighting conditions. This increases the number of iterations and may lead ASM fail to detect the exact eye shape.

In the improved ASM, training stage, 1-D profile is replaced by 2-D profile called sub-image. The sub-image lies in two dimensions and outnumbers the pixels contained in a 1-D profile. By using the \( M \) eye images in database, we follow the same strategy of the standard ASM for collecting a set of sub-images and calculating the
Fig. 7 Improved searching operation. (a) Sub-image collected at the first point. (b) Corresponding first 16 eigenvectors corresponding mean sub-image at each shape point as practically shown in Fig. 7(a) for the first eye shape point.

In detection stage, searching operation, we also follow the strategy of creating searching lines as shown in Fig. 2(c). A sub-image along the searching direction that best matches the mean sub-image is considered as the target sub-image and its location is the new location of the processed eye shape point.

The sub-image approach imposes two conditions; first, the searching operation must be able to measure the similarity between the sub-images along the searching line and those collected from the database. This property ensures distinguishing the eye boundary from the skin and eyeball areas. Second, the searching operation must have an ability of generalization. Generalization means that the searching operation can predict/detect the target sub-image (location) even if it is not presented by the collected sub-images from the database. This property ensures more flexibility to deal with various eye profiles and lighting conditions.

In order to satisfy the above two conditions, PCA is used to model the variation of pixel values exhibited in the collected sub-images. For each eye shape point, a mean sub-image and an eigenvector matrix are computed as shown in Fig. 7(a-b) and in the same manner stated in subsection 3.1.2.

Generalization is achieved by the elliptical distribution formed by the first few eigenvectors. Any sub-image belongs to this ellipse can be considered as a new representation of the target sub-image. On the other hand, distinguishing the eye boundary from the skin and eyeball areas along a searching line created at \( i^{th} \) point is guaranteed by measuring the distance/correlation between decomposition vector of the mean sub-image of the \( i^{th} \) point and those of sub-images along the searching line [16].

One may ask why an eigenspace is created for each eye shape point instead of one eigenspace for all points?. Each eye shape point possesses a specific pattern of sub-images. Logically and according to the conducted experiments, an eigenvector space tend to house a cluster such as that in Fig. 5. Therefore, if sub-images of all eye shape points are combined to create one eigenspace, the tight elliptical distribution of the corresponding cluster generated is vanished and a random sphere is obtained. Consequently, non-linearity is increased and PCA loses the generalization property.

Figure 8 shows the results of the improved searching operation based on the initial estimation of the eye shapes in Fig. 6(b).

4. Experimental Results
The improved and standard ASMs have been implemented in Visual Studio C++ 2008 environment. Since the aim of this paper is to detect the eye shape in an eye image, we used an eye detection technique based on Haar-Like Feature method to obtain the eye images [17].

The number of eye images collected for the experiments is 150; 100 images are used for eye database and 50 images are used for evaluating the performance of the standard and improved ASMs. The shapes of 150 images have been labeled manually using 16 points. Two points, \( 1^a \) and \( 9^b \), represent the eye corners and the remaining points are equally divided to represent the upper and lower eyelids. The images in database are carefully selected to satisfy two conditions. 1) Slight differences in rotation, scale and eye states (nearly closed, open and in-between) to strong up the initial estimate decision. 2) The images are taken under different lighting conditions to make the searching operation more effective using sub-images strategy. The geometrical structure of the eye has been analyzed by the first seven eigenvectors of the matrix \( \Psi \).

An eigenvector spaces has been created for each shape point using the corresponding collected sub-images form database. The size of each sub-image is 7x7 pixels. The
first seven eigenvectors were found very appropriate to
describe the variation of pixel values exhibited in each set
of sub-images. Standard ASM setup only differs in the use
of a 1-D profile which contains 7 pixels.

Equation (10) is used to measure the best correlation
between the decomposition vectors of a testing eye image
\( D_t \) and those in eye image database in order to select the
proper initial estimation.

\[
\frac{D_i \cdot D_j}{\|D_i\| \cdot \|D_j\|}, \quad 0 \ll i \ll M
\]  

(10)

The same equation is applied to detect the new location
of an eye shape point by measuring the best correlation
between decomposition vectors of the sub-images along
the searching line and the mean sub-image.

Equation (11) is used to measure the match error \( Mrr \)
between a manually labeled eye shape, exact shape, \( X_e \) of a
testing image and the detected eye shape \( X_d \) by the
standard or improved ASM. Finally, each method
converges when the difference between the eye shapes of
last two iterations is less than \( \alpha = 0.005 \).

\[
Mrr\% = \left[1 - \frac{X_e \cdot X_d}{\|X_e\| \cdot \|X_d\|}\right] \times 100
\]  

(11)

Figure 9 shows time consumption as a function of the
number of eye images in database. PCA consumes time to
recognize and understand the geometrical structure of a
given image. The relationship is linear because the eye
images are represented by a fixed number of eigenvectors.

Figure 10 illustrates the relationship between the
number of iterations and the match error. The curves
demonstrate the development of the eye shape during the
searching operation. The beginning of each curve
highlights the match error of the initial eye shape
estimation. The end points represent the match error of the
detected eye shape.

Figure 11 represents the change of the match error with
respect to the detection time. The curves can be utilized to
evaluate the behavior of searching operation. The curve of
the proposed method starts after around 10 msec that
consumed by recognition operation.

5. Discussion

PCA is usually used for object recognition, especially for
recognizing human faces [18]. In this paper, PCA is
employed for discriminating the eye profile at different
parameters. Eye is surrounded by skin whereas human face
is surrounded with various backgrounds. Therefore, the
change of the eye profile is the dominant feature in the first
few eigenvectors whereas in a face image, there are many
features (nose, hair, two eyes states, background…,etc)
that may dominant the first eigenvectors behavior
frequently. Thus, we said earlier that PCA is used to
understand the geometrical structure of an eye state instead
of recognizing the eye of a specific person.

In the proposed method, the initial estimation
becomes more robust to process the change of the eye
profile. According to the conducted experiments, 100 eye
shapes have been used as a training set. These eye shapes

Figure 9 The relationship between time of recognition and
number of images in database

Figure 10 The relationship between match error and number
of iterations

Figure 11 The relationship between match error and time
consumption

have different parameters (scale, rotation and eye state).
Based on object recognition and according to the eye
profile in the testing image, one of these shapes is used as
an initial estimate instead of always using the mean shape.

From another point of view, Fig. 9 shows that PCA
consumes time to understand the eye geometrical structure.
Thus, increasing the number of eye images in database
must be taken into account especially in real time process
such as tracking the eye shape in a sequence of images.
100 eye images consumes around 10msec which not affect
the real time system comparing to the advantage gained by
the improved initial estimation. Figure 10 proves this point by showing a big difference in the match error at the beginning of the curves. The standard ASM consumes a number of iterations to reach the initial match error of the improved ASM. This reflects significantly on time consumption as shown in Fig.11.

The sub-image approach is used to reduce the number of iterations and increase the accuracy of the detected eye shape. The 1-D mean profile used by the standard ASM consists of 7 pixels and the number of pixels contained in a sub-image is 7x7 pixels. The sub-images are represented by only 7 projections in the eigenvector space. This means that the consumed time in detecting the new location of an eye shape point in both methods is same. On the other hand, two advantages are gained by using the improved searching operation. First, the sub image contains a larger number of pixels. Second, the correlation measurement is more robust in the eigenvector space. Figure 11 highlights that the searching operation becomes smoother and more stable compared to the standard ASM. Consequently, time consumption is reduced and high accuracy is achieved.

However, dealing with complex eye states such as nearly closed is still a problem for the proposed method. A suggested solution is to increase the number of images in database that represents more different eye states. This will lead to increase the recognition time accordingly. In order to preserve the same number of images in database and contain more different eye states, the pupil of the images in database is approximately filtered out. Thus, the eye structure will dominate the first eigenvectors more efficiently and hence the number of images taken to describe an eye state in database will be reduced as well.

In addition, the improved searching operation based-sub-image has increased the robustness of ASM to deal with different lighting conditions. On the other hand, the performance is negatively influenced when a change occurs far from the middle range of brightness. In order to overcome this problem, ASM can be charged with various lighting condition databases. Based on the lighting conditions in a testing image, a database will be selected and corresponding eigenvectors of object recognition and sub-images will be used to detect the eye shape in the testing image. This work is left to future.

6. Conclusion

In this paper, some steps to improve the performance of Active Shape Model in detecting the eye shape are proposed. Principal Component Analysis (PCA) is employed to understand the geometry of an eye image and then to select a proper eye shape from the training set. This step has increased the accuracy of the initial estimation around 50% compared to the standard ASM. PCA is also used to strong up the behavior of searching operation using the sub-image approach. This step has reduced the number of iterations and consequently the detection time to around 60 msec whereas the standard ASM needs more than 200 msec to reach the same value. Therefore, the proposed method is very appropriate to be used in real time systems.

References