A Preprocessing Method for Active Appearance Models

Sung Joo Lee[1], Kang Ryoung Park[2], Jaihie Kim[1]
[1] School of Electrical and Electronic Engineering, Yonsei University,
Biometrics Engineering Research Center
#401, Industry-University Research Center, Yonsei University 134
Shinchon-dong, Seodaemun-gu, Seoul 120-749, South Korea

[2]Department of Electronics Engineering, Dongguk University,
Biometrics Engineering Research Center
26, Pil-dong 3-ga, Jung-gu, Seoul 100-715, South Korea
E-mail: [1]{sungjoo, jhkim}@yonsei.ac.kr, [2] parkgr@dongguk.edu

Abstract: Active Appearance Models (AAMs) are parametric facial models which have been widely used for facial features localization. In order to localize facial features, AAMs used an optimization method which finds local minima. As a result, AAMs cannot always localize facial features exactly when initial locations of facial features are far from the ground truth locations of them. To solve this problem, we propose a simple preprocessing method which finds initial locations of facial features such as eyes, nostrils, and lip corners. Experimental results showed that the proposed method was robust to moderate pose variation, illumination condition changes, complex background, and glasses wearers.

1. Introduction

Active Appearance Models (AAMs) are a kind of parametric facial models which have been widely used for facial features localization[1][2]. Generally, the parametric facial model used in AAMs is divided into parametric shape model and appearance model. In order to localize facial features, AAMs used an optimization method to minimize the warped input face image and the synthesized face image by using the parametric model. However, the optimization method cannot find global minima but find local minima. As a result, AAMs cannot always localize facial features exactly when initial locations of facial features are far from the ground truth locations of them as shown in Fig 1. To solve this problem, we propose a simple preprocessing method which finds initial locations of facial features such as eyes, nostrils, and lip corners. From input image, a face-only image was localized by the adaboost algorithm. Then, the face-only image was binarized, morphologied, and labeled to find the initial faical features. Experimental results showed that the proposed method was robust to moderate pose variation, illumination condition changes, complex background, and different camera and glasses wearers.

2. Backgournds

In order to describe human facial variations, AAMs have statistical facial shape and appearance models using principle component analysis (PCA). A facial shape $\mathbf{s}$ that consists of k vertices, $(x_1, y_1), (x_2, y_2), \cdots, (x_k, y_k)$, is represented by a linear combination of the mean facial shape $\mathbf{s}_0$ and n facial shape variation vectors $\mathbf{s}_i$:

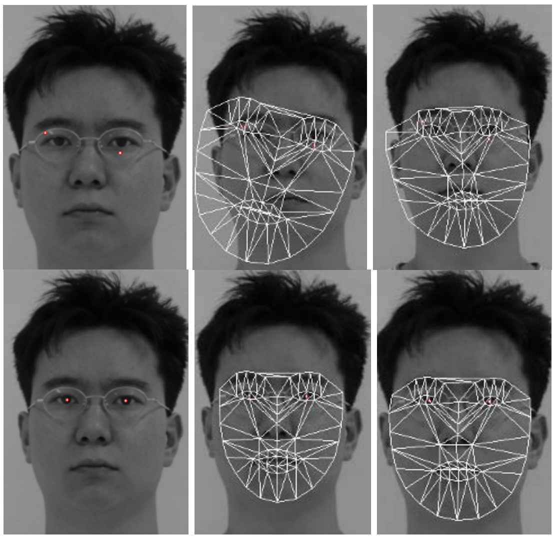

Figure 1. Two different initial facial feature points(first column), corresponding initial shape(second column), and fitted facial shape(third column)

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} \alpha_i \mathbf{s}_i \tag{1}$$

where, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]$ refers to the shape parameter vector and $\mathbf{s} = [x_1, y_1, \cdots, x_k, y_k]$. Likewise, the facial appearance $\mathbf{A}$ is modeled by the linear combination of the mean facial appearance $\mathbf{A}_0$ and m facial appearance variation vectors $\mathbf{A}_i$:

$$\mathbf{A} = \mathbf{A}_0 + \sum_{i=1}^{m} \beta_i \mathbf{A}_i \tag{2}$$

where, $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_m]$ presents the appearance parameter vector. $\mathbf{s}_i$ and $\mathbf{A}_i$ were obtained from the training facial images using PCA. In addition, training facial appearances were normalized to get the shape-free facial appearances before using PCA [1][2]. Then, the appearance model for the particular image pixel can be written as:

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \beta_i A_i(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbf{p}(\mathbf{s}_0) \quad (3)$$

where, $\mathbf{p}(\mathbf{s}_0)$ denotes the set of pixels $\mathbf{x} = (x, y)^T$ that lie inside the mean facial shape $\mathbf{s}_0$ and $A(\mathbf{x})$ means the intensity value of the facial appearance A at point $\mathbf{x}$. The goal of AAM fitting methods is to find the shape and appearance parameter vectors that minimize errors between the synthesized face image and the warped input face image. The objective function is as follows:

$$\sum_{\mathbf{x} \in \mathbf{p}(\mathbf{s}_0)} [A_0(\mathbf{x}) + \sum_{i=1}^{m} \beta_i A_i(\mathbf{x}) - I(W(\mathbf{x};\boldsymbol{\alpha}))]^2 \quad (4)$$

where, W represents the warping function to change the point location from point $\mathbf{x}$ in the input face image coordinate to point $W(\mathbf{x};\boldsymbol{\alpha})$ in the synthesized image coordinate[2].

The the Simultaneous Inverse Compositional(SIC) algorithm minimizes Eq. (4) by performing a Gauss-Newton gradient descent optimization simultaneously on the shape parameter $\boldsymbol{\alpha}$ and appearance parameter $\boldsymbol{\beta}$ [6][7]. The algorithm iteratively minimizes:

$$\sum_{\mathbf{x} \in \mathbf{p}(\mathbf{s}_0)} [A_0(W(\mathbf{x};\Delta\boldsymbol{\alpha})) + \sum_{i=1}^{m} (\beta_i + \Delta\beta_i) A_i(W(\mathbf{x};\Delta\boldsymbol{\alpha})) - I(W(\mathbf{x};\boldsymbol{\alpha}))]^2 \quad (5)$$

with respect to $\Delta\boldsymbol{\alpha}$, $\Delta\boldsymbol{\beta} = [\Delta\beta_1, \Delta\beta_2, \cdots, \Delta\beta_m]$. If we denote the combined parameter vector $\mathbf{c}$ and update of $\mathbf{c}$ as:

$$\mathbf{c} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}, \ \Delta\mathbf{c} = \begin{pmatrix} \Delta\boldsymbol{\alpha} \\ \Delta\boldsymbol{\beta} \end{pmatrix} \quad (6)$$

then, the $\Delta\mathbf{c}$ can be found as follows:

$$\Delta\mathbf{c} = -H^{-1} \sum_{\mathbf{x}} \mathbf{SD}^T(\mathbf{x}) E_{app}(\mathbf{x}) \quad (7)$$

where, the steepest descent image $\mathbf{SD}$, the hessian H, and the appearance error $E_{app}(\mathbf{x})$ were calculated as follows:

$$\mathbf{SD}\ (\mathbf{x}) = \left(\nabla\mathbf{A}\frac{\partial W}{\partial \alpha_1}, \cdots, \nabla\mathbf{A}\frac{\partial W}{\partial \alpha_n}, \mathbf{A}_1(\mathbf{x}), \cdots, \mathbf{A}_m(\mathbf{x}) \right) \quad (8)$$

$$H^{-1} = \sum_{\mathbf{x}} \mathbf{SD}^T(\mathbf{x})\mathbf{SD}\ (\mathbf{x}) \quad (9)$$

$$E_{app}(\mathbf{x}) = I(W(\mathbf{x};\boldsymbol{\alpha})) - \left[A_0(\mathbf{x}) + \sum_{i=1}^{m} \beta_i A_i(\mathbf{x}) \right] \quad (10)$$

A detail description on this algorithm can be found in [6][7].

Besides SIC algorithm, many algorithms have been proposed [1][2]. Unfortunately, these algorithms cannot always find global minima. As a result, these algorithms fall into local minima occasionally when initial shape parameter vector is far from the ground truth of it. Therefore, initial shape parameter vector should approximate to the ground truth of it. To solve this problem, we propose a simple preprocessing method which find initial locations of facial features such as eyes, nostrils, and lip corners.

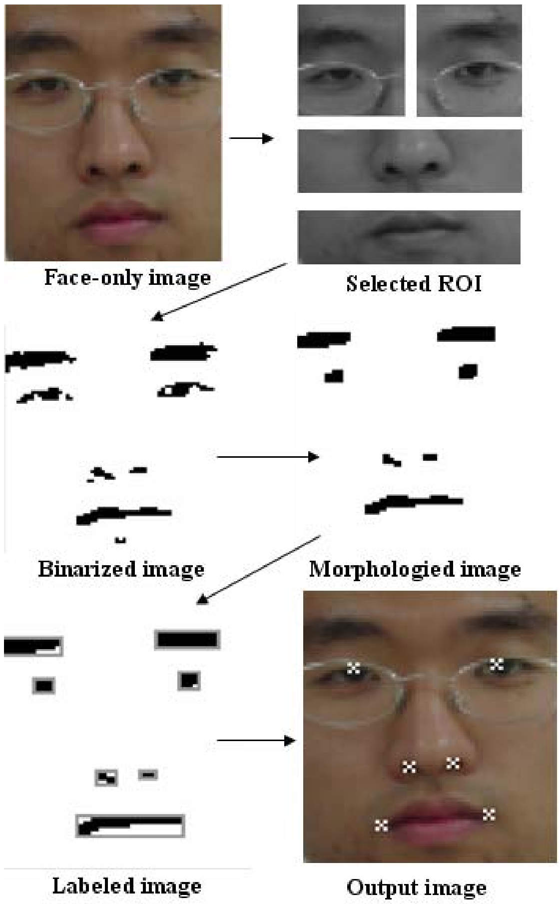

Figure 2. The flow chart of the proposed method.

3. Proposed method

The overall procedure of the proposed method consists of face detection, Region Of Interest (ROI) division, binarization, morphology, and labeling as shown in Fig 2.

First stage is face detection. Usually, in practical case, a complex background exists when we capture the face image. It is very difficult to find the initial facial features when the complex background exists because the background sometimes have similar color, intensity, and shape properties of the initial facial features. Therefore, we used the Adaboost algorithm which has been known as reliable face detector to remove complex background as shown in figure 3 [3].

Second stage is ROI division which is needed for reducing ROI and focusing on each facial feature. After converting color image to gray image, the ROI divided based on assumptions that symmetric eyes may exists in upper half of the face, nostrils may exists in range of half to third quarter of the face, and lip corners may exists in the other fourth quarter of the face. These assumptions can deal with moderate pose variation and were validated empirically.

Figure 3. Face detection result. Original image (left), Face-only image (right)

Figure 4. Binarization result. Original images under different illumination conditions (left column), corresponding binarized images (right column).

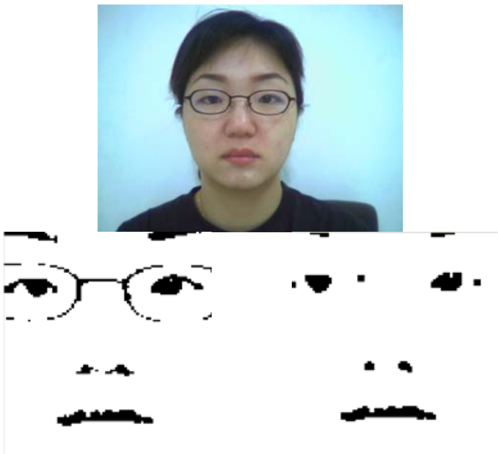

Figure 5. The result of morphology: original image (top row), binarized image (bottom left), and morphologied image (bottom right).

Third stage is binarization based on intensity properties of the initial facial features. Generally, eyes, nostrils, and lip corners are relatively dark in the corresponding ROI. Therefore, the initial facial features can be binarized by using the p-tile thresholding method [4]. As shown in figure 4, this approach is robust to illumination condition changes because, in small ROI, brightness variations caused by illumination changes are small and the p-tile thresholding method determine the threshold value by using relative brightness information (percentage) instead of a constant threshold value. In this paper, the threshold value p is set to 10% for eyes localization, 2% for nostrils localization, and 7% for mouth corner localization respectively. These values are determined empirically.

Fourth stage is morphology to remove unnecessary noise such as a glasses frame, eyelids, and so on. As shown in figure 5, a black glasses frame can be removed by using morphology. The opening operator with typical 3 by 3 square and 2 by 2 rectangular structuring elements were used for eyes and the other facial features respectively.

Final stage is labeling to localize the initial facial features. Connect components were labeled by using open source [5]. Then, position, size and shape constraints are used to localize the initial facial features.

4. Experimental results

The proposed method was evaluated to moderate pose variations, different illumination conditions, complex background, different camera and glasses wearers. The results were successful as shown in Figures 6 to 11. However, the proposed method can not find a nostril when there is a large head rotation because the nostril is occluded as shown in figure 12.

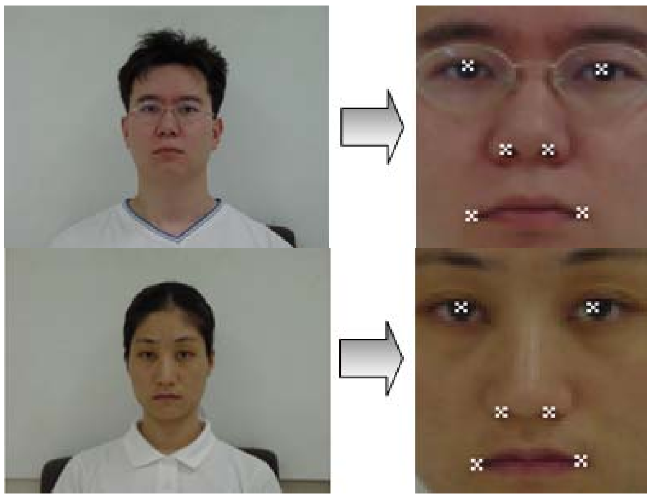

Figure 6. An image in normal pose with a digital camera

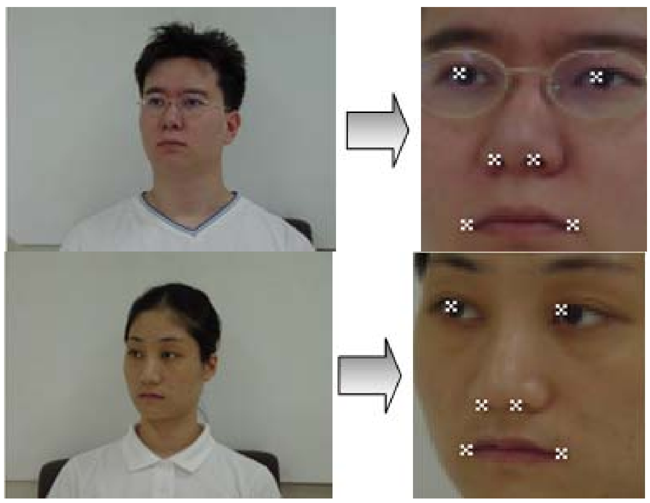

Figure 7. An head rotated image with a digital camera.

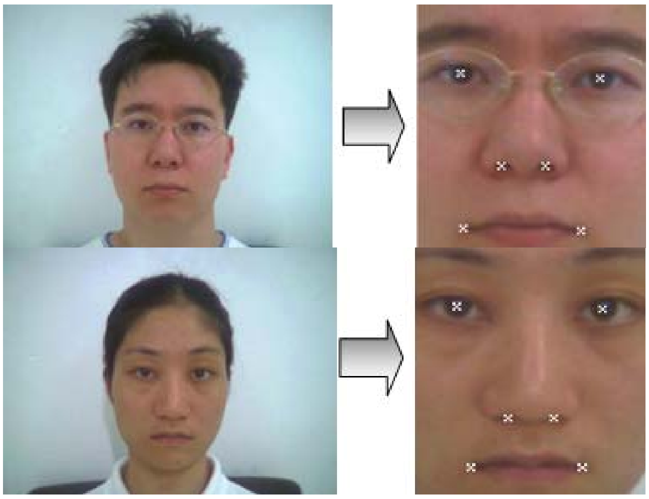

Figure 8. An image with a mobile camera

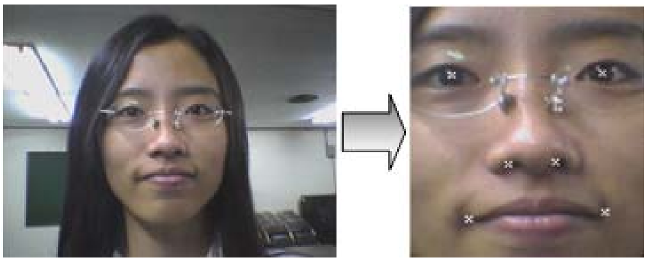

Figure 9. An image with a complex background

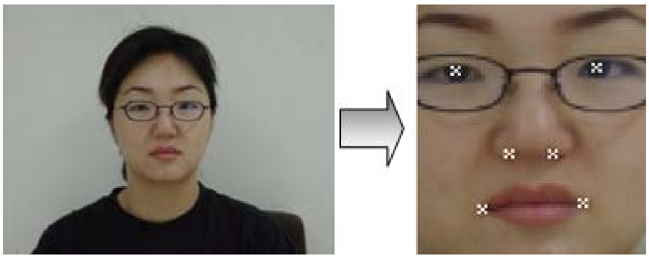

Figure 10. An image with a black glasses frame

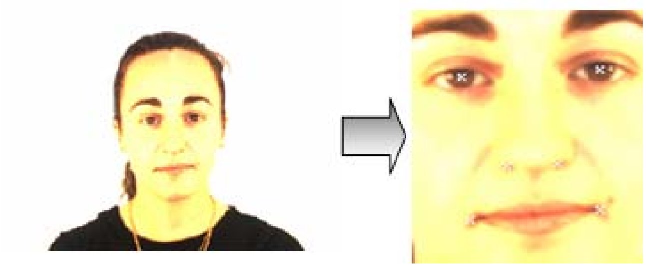

Figure 11. An image with large illumination.

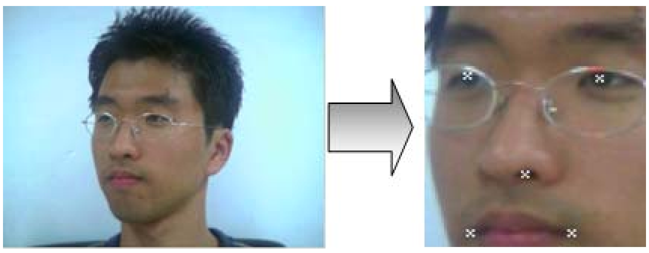

Figure 12. An image with large rotation.

5. Conclusion

In this paper, we proposed a simple preprocessing method, which can find the initial facial feature points in order to reduce AAMs fitting error probability caused by local minima. Experimental results showed that the proposed method was robust to moderate pose variation, illumination condition changes, complex background, different camera and glasses wearers. As a future work, we will combine the proposed method with AAMs and measure quantitative errors on various imaging conditions.

Acknowledgements

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University. (R112002105070030(2008))

References

[1] T. Cootes, G. Edwards, C. Taylor, "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23 no. 6, pp. 681–685, 2001.

[2] I. Matthews, S. Baker, "Active Appearance Models revisited", International Journal of Computer Vision vol. 60, no. 2, pp. 135–164, 2004.

[3] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Proc. CVPR, vol.1, pp.511-518, 2001.

[4] Rames Jain, Rangachar Kasturi, Brian G. Schunck, "Machine Viosion", McGrawHill, Inc.,1995.

[5] http://opencvlibrary.sourceforge.net/cvBlobsLib, accessed on May 23, 2008.

[6] Baker, S., Gross, R., Matthews, I., 2003. Lucas-Kanade 20 Years on: A Unifying Framework: Part 3, Technical Report CMU-RI-TR-03-35, Carnegie Mellon University Robotics Institute.

[7] Gross, R., Matthews, I., Baker, S., 2005. Generic vs. person specific active appearance models, Image and Vision Computing, 23(11), pp. 1080–1093.