조명 영상 합성을 통한 AAM 피팅 성능 개선

이형수[O] 김대진
포항공과대학교 컴퓨터공학과 지능형미디어 연구실
{sooz[O], dkim}@postech.ac.kr

Fitting Enhancement of AAM Using Synthesized Illumination Images

Hyung-Soo Lee[O] Daijin Kim
I.M. Lab., Dept. of Computer Science & Engineering, POSTECH

Abstract

Active Appearance Model is a well-known model that can represent a non-rigid object effectively. However, since it uses the fixed appearance model, the fitting results are often unsatisfactory when the imaging condition of the target image is different from that of training images. To alleviate this problem, incremental AAM was proposed which updates its appearance bases in an on-line manner. However, it cannot deal with the sudden changes of illumination. To overcome this, we propose a novel scheme to update the appearance bases. When a new person appears in the input image, we synthesize illuminated images of that person and update the appearance bases of AAM using it. Since we update the appearance bases using synthesized illuminated images in advance, the AAM can fit their model to a target image well when the illumination changes drastically. The experimental results show that our proposed algorithm improves the fitting performance over both the incremental AAM and the original AAM.

1. Introduction

Active Appearance Model (AAM) is a generative model that allows both the shape and appearance variations [1]. These variations are represented by linear models such as Principal Component Analysis (PCA), which finds a subspace reserving maximum variance of given data. The AAM has been widely used and has many application areas such as face modeling, medical image modeling and so on.

However, since the AAM uses the fixed appearance model, the fitting results are often unsatisfactory especially when the illumination condition of the target image is far different from that of training images which are used to learn the appearance model. This problem can be solved by collecting a large number of training images that contain every possible illumination conditions, but collecting such training images is impossible. To alleviate this problem, we proposed to use adaptive linear appearance model that update its appearance bases using the incremental PCA [2]. However the update of appearance bases using ill-fitted images can worse their fitting performance of the AAM than that of original AAM. Hence we used modified adaptive observation model (AOM)[3] as a measure to determine whether to update the appearance bases or not, when a new fitting result is given. By this scheme, we first fit the AAM to the input image, and then determine the goodness of the fitting result by computing the percentage of outlier pixels. If the fitting result is good, the AOM parameters are updated and the new appearance bases of the AAM are computed using the incremental PCA. Then, the updated AOM and AAM are used for the next frame image. This algorithm works well under the gradual change of illumination. However, when the illumination condition changes drastically, the AOM judges the entire warped pixel as outlier pixels. In a consequence, neither the AOM nor the AAM is updated and the fitting performance is not improved.

To overcome the drawback of the incremental AAM, we propose a novel scheme to update the appearance bases. The drawback of the previous work is that the AOM cannot adapt to the rapid change of illumination, as a result, no update of appearance bases is take place. Therefore, instead of using AOM, when a new person appears in the input image, we synthesize illuminated face images of that person and update the appearance bases using it. By doing this, the appearance bases can fit to the input image accurately. In addition, even when the illumination condition changes drastically, the AAM fit to the input image well by the virtue of the appearance bases which are updated using the synthesized illuminated face images. The advantage of the proposed algorithm over the incremental AAM is that since we update the appearance bases only once when a new person appears, we do not need to determine the goodness of the fitting result during the image sequences. The experimental results show that our proposed algorithm improves the fitting performance over both the incremental AAM and original AAM.

2. Theoretical backgrounds

2.1 Active appearance models

In 2D AAM[1], the 2D shape is represented by a triangulated 2D mesh with l vertices, which correspond to the salient points of the object. Mathematically, the shape vector $\mathbf{s}$ consists of the 2D coordinate of the vertices that make up the mesh as $\mathbf{s} = (x_1, y_1, \dots, x_l, y_l)^t$ and shape variation is expressed by a linear combination of a mean shape $\mathbf{s}_0$ and n shape bases $\mathbf{s}_i$ as

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

where p_i are the shape parameters. A standard approach to compute the linear shape model is to apply the principal component analysis (PCA) to a set of shape vectors that are gathered from the manually landmarked training images and aligned using Procrustes analysis, where the ith shape basis $\mathbf{s}_i$ is the ith eigenvector that corresponds to the ith largest eigenvalue.

Once a mean shape $\mathbf{s}_0$ is obtained, the training images are warped to the mean shape using the piece-wise affine warp that is defined between the corresponding triangles in the landmarked shape of the training images and the mean shape. Then, we can define the appearance as a shape normalized image $A(\mathbf{x})$ over the pixels $\mathbf{x}$ that belong to the inside of the $\mathbf{s}_0$. The appearance variation is expressed by a linear combination of a mean appearance $A_0(\mathbf{x})$ and m appearance bases $A_i(\mathbf{x})$ as

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i A_i(\mathbf{x}), \tag{2}$$

where α_i are the appearance parameters. The appearance model is computed from the manually landmarked training images by collecting the shape normalized images and applying PCA to them, where the ith appearance basis image $A_i(\mathbf{x})$ is the ith eigenvector that corresponds to the ith largest eigenvalue.

2.2 Incremental principal component analysis

To incrementally update the appearance bases such that the updated linear appearance model can represent a new appearance data, the traditional PCA requires to keep all the training images, which is very inefficient. Instead we use incremental subspace learning algorithm[4] that is more efficient than traditional PCA.

Suppose that a set of d-dimensional data vectors is $D = \{\mathbf{d}_1, \dots, \mathbf{d}_N\}$. The eigenspace of the data set can be obtained by solving the singular value decomposition (SVD) of the covariance matrix $\mathbf{C}$. Then, the given data set can be represented by $k(< d)$ -dimensional coefficient vectors $\mathbf{a}_i$ by projecting the data vector $\mathbf{d}_i$ to a subspace spanned by k eigenvectors corresponding to k largest eigenvalues. For the ease of the explanation, we will use a matrix $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_k] \in \mathbb{R}^{d \times k}$ that contains the k eigenvectors and a diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{k \times k}$ that contains k large eigenvalues as the diagonal elements in the descending order.

When a new data vector $\mathbf{d}_{N+1}$ is given, the incremental PCA updates the mean and the basis vector as follows. Since the total amount of the data is changed, we should update the mean and the basis vector to represent the data including a new data. The mean is updated as

$$\mathbf{d}' = \frac{1}{N+1}(N\mathbf{d} + \mathbf{d}_{N+1}). \tag{3}$$

Then, the orthogonal residual vector $\mathbf{b}_{N+1}$ is computed as

$$\mathbf{b}_{N+1} = (\mathbf{U}\mathbf{a}_{N+1} + \mathbf{d}) - \mathbf{d}_{N+1}. \tag{4}$$

Let a normalized vector be

$$\mathbf{b}_{N+1} = \frac{\mathbf{b}_{N+1}}{\mathrm{P}\mathbf{b}_{N+1}\mathrm{P}_2}. \tag{5}$$

We acquire the new basis set $\mathbf{U}'$ by rotating the basis set $[\mathbf{U} \quad \mathbf{b}_{N+1}]$ so that the i-th basis of the new basis represents the i-th largest maximal variance as the

$$\mathbf{U}' = [\mathbf{U} \quad \mathbf{b}_{N+1}]\ \mathbf{R}. \tag{6}$$

The rotation matrix can be obtained by solving SVD for $\mathbf{D}$ matrix:

$$\mathbf{D}\ \mathbf{R} = \mathbf{R}\ \mathbf{\Lambda}', \tag{7}$$

when we compose $\mathbf{D} \in \mathbb{R}^{(k+1)x(k+1)}$ as

$$\mathbf{D} = \frac{N}{N+1}\begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \frac{N}{(N+1)^2}\begin{bmatrix} \mathbf{a}\mathbf{a}^T & \boldsymbol{\beta}\mathbf{a} \\ \boldsymbol{\beta}\mathbf{a}^T & \boldsymbol{\beta}^2 \end{bmatrix}, \tag{8}$$

where $\boldsymbol{\beta} = \mathbf{b}_{N+1}^T(\mathbf{d}_{N+1} - \bar{\mathbf{d}})$ and $\mathbf{a} = \mathbf{U}^T(\mathbf{d}_{N+1} - \bar{\mathbf{d}})$.

2.3 Bilinear Model

The bilinear model[5] is a two-factor model that separates the observations into two factors such as style and content. When we see a character, we separate it as a font (style) and a meaning (content), which are independent factors that represent the character. The bilinear model is categorized into two types: symmetric model and asymmetric model. In symmetric model, the bilinear model interacts with the style and content using an interaction matrix that makes them independent. We used the symmetric model for the synthesis of

illuminated face images.
A symmetric bilinear model represents the observation vector y as

$$y = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} a_i b_j, \tag{9}$$

where w is a basis vector which interacts with style factor a and content vector b and the size of the two vectors is K. To use the symmetric bilinear model, we need to learn the interaction basis vector w. Assume that we have $S \times C$ training samples and we build the observation matrix Y by stacking them:

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1C} \\ \vdots & \ddots & \vdots \\ y_{S1} & \cdots & y_{SC} \end{pmatrix}, Y^{VT} = \begin{pmatrix} y_{11} & \cdots & y_{1S} \\ \vdots & \ddots & \vdots \\ y_{C1} & \cdots & y_{CS} \end{pmatrix}, \tag{10}$$

where the superscript VT means vector transpose and each element y_{ij} is a K-dimensional observation vector. The observation matrix Y has a size of SK $\times$ C. Then, the symmetric bilinear model can be represented in a compact form as

$$Y = (W^{VT}A)^{VT} B or Y^{VT} = (WB)^{VT} A, \tag{11}$$

where A and B represent the stacked style and content factor matrices whose size are $I \times S$ and $J \times C$, respectively:

$$A = (a_1, \cdots, a_S), B = (b_1, \cdots, b_C) \tag{12}$$

and W is the stacked interaction matrix.
Usually, the optimal style and content matrices A and B are estimated by an iterative computation using singular value decomposition(SVD), because it tends to get the global and non-localized features.
As a result of bilinear model learning, we have the interaction matrix W, the style parameter matrix A, the content parameter matrix B for the symmetric bilinear model. Given a single test vector, we can obtain the estimate of the style vector a and the content vector b with respect to the interaction matrix W. The detailed algorithm for obtaining the style and content factors a and b of a test vector y is given in[5].

3. Illumination image synthesis and its application to incremental AAM

3.1 Illumination image synthesis

Many researchers have been tried to synthesize new illuminated image from input image which is captured under arbitrary illumination condition. Sim and Kanade[6] proposed a model and example based method to synthesize a new illuminated image. Belhumeur and Kriegman[7] introduced the illumination convex cone. They argued that the images under all possible illumination conditions built a convex cone in the image space and the reconstructed shape and albedo of the face from a small number of samples served as a generative model for synthesizing images of the face under novel poses and illumination conditions. However this algorithm requires at least three images of the same face taken under different lighting conditions. Shashua[8] introduced the quotient image that uses class-based re-rendering and recognition with varying illuminations. They defined an illumination invariant signature image that enables an analytic generation of images with varying illuminations. However, their approach might fail in obtaining the illumination invariant feature when the input image has a shadow.
We propose a novel illumination image synthesis method which adopts ratio image concept into bilinear model framework. First, we assume that the face has the Lambertian surface: a face image can be represented by the product of the albedo, the surface normal, and a light source. Then the intensity of a pixel at the position (x, y) in the image is represented as

$$I(x, y) = \rho(x, y) n(x, y)^T \cdot s, \tag{13}$$

where, $\rho(x, y)$ is the albedo of pixel (x, y), $n(x, y)$ is the surface normal, and s is the point light source, respectively. In addition, we warp the input face image into a predefined mean shape by means of AAM. Therefore we also assume that all the people have the same surface normal. Denote the input face image under an arbitrary illumination s as $I_{in}^s = \rho_{in}(x, y) n(x, y)^T \cdot s_s$. Ratio image of input face between two different illumination A, B is defined as follows:

$$R_{in}^{AB}(x, y) = \frac{I_{in}^B(x,y)}{I_{in}^A(x,y)}. \tag{14}$$

From Eq. (13), we have

$$R_{in}^{AB}(x, y) = \frac{\rho_{in}(x,y) n(x,y)^T \cdot s_B}{\rho_{in}(x,y) n(x,y)^T \cdot s_A} = \frac{n(x,y)^T \cdot s_B}{n(x,y)^T \cdot s_A}. \tag{15}$$

Here, we can see that surface normal and light source determine the ratio image. Moreover, since we assume that all the people have the same surface normal, ratio image between two different illumination is the same for all the people. From now, we represent the ratio image between two illumination A and B as R^{AB} by dropping person-specific term from Eq. (14).
Our goal is to synthesize a face image I_{in}^t under a novel light source t, given input face image I_{in}^s under light source s and a reference image I_{ref}^t under target light source t. From the ratio image,

$$R^{st}(x,y) = \frac{I_{in}^{t}(x,y)}{I_{in}^{s}(x,y)} = \frac{I_{ref}^{t}(x,y)}{I_{ref}^{s}(x,y)}. \quad (16)$$

The target image I_{in}^{t} can be calculated if we can estimate $I_{ref}^{s}(x,y)$ which is the reference image under input light source s. We adopt symmetric bilinear model for estimating $I_{ref}^{s}(x,y)$ and treat the facial identity and the lighting as a content factor and a style factor of bilinear model, respectively. First we factorize the input face image into the identity factor and the lighting factor. Then we can synthesize the reference face image under the input lighting condition by multiplying the identity factor of the reference face image, the lighting factor of the input face image, and the interaction matrix of bilinear model. Here, we used mean of the training face images as the reference face image.
The detailed explanation of the overall procedure of the proposed illumination image synthesis method is given below.

- Factorize the input face image into identity factor and lighting factor as $I_{in}^{s} = a_{in}Wb_{s}$, where a_{in} is the identity factor, b_{s} is the lighting factor, and W is interaction matrix, respectively.

- Obtain the reference face image under the input lighting condition $I_{ref}^{s} = a_{ref}Wb_{s}$ using the identity factor of the reference image a_{ref} and the lighting factor of the input image b_{s}.

- Obtain the ratio image between lighting s and t: $R^{st}(x,y) = \frac{I_{ref}^{t}(x,y)}{I_{ref}^{s}(x,y)}$.

- Compute the target face image $I_{in}^{t}(x,y) = R^{st}(x,y)I_{in}^{s}(x,y)$.

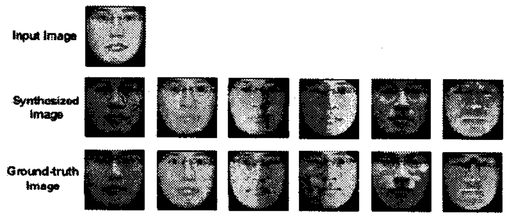

Fig. 1. The reslt of illumination image synthesis

Fig. 1. shows the result of illumination image synthesis. We can see that the synthesized images (row 2) have almost identical illumination condition with the ground-truth images (row 3) and the identity of the input face is not changed. Hence we can apply the proposed illumination image synthesis method to incremental AAM.

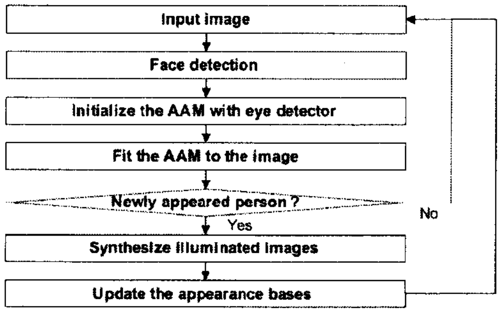

Fig. 2. The procedure of the incremental AAM using synthesized illumination images

4. Incremental AAM using synthesized illumination face images

To improve the fitting performance of AAM, incremental AAM updates the appearance bases as the imaging condition changes. However, when the illumination condition changes drastically, the AOM judges the entire warped pixel as outlier pixels. In a consequence, neither the AOM nor the AAM is not updated and the fitting performance is not improved. To solve this problem, we first synthesize illuminated face images and update the appearance bases using it. As a result, we expect the fitting performance of AAM to improve.
Fig. 2. shows the procedure of the incremental AAM using synthesized illuminated images. For a new image, the AAM is initialized using a face and eye detector and fitted to the input image. Next, the algorithm determines whether the fitted face is a newly appeared person using the face tracker. If the fitted face is a newly appeared person, the illuminated face images are synthesized as explained in section 3 and the appearance bases of AAM is updated using theses images. Then, the updated AAM is used for the next input image.

5. Experimental results

5.1 Data setup

For constructing active appearance model, we used the face image database that were gathered from 28 people in our lab. For each person, 6 images are registered in the database and the images contain four facial expression variations(three neutral images, a happy image, a surprise image, and a angry image) at frontal view. Therefore there are 168 images and we manually landmarked all the images. Then we constructed the model and it is built using 37 appearance bases and 28 shape bases. Each number of bases is selected to account for 95% shape and appearance variations.
For training the bilinear model, we also gathered

another face image database which consists of 15 people. For each person, 6 images under different lighting directions are registered. Therefore there are 90 images and we manually landmarked all the images. Then we warped each face image into the mean shape of the AAM and we used this warped face image to train the bilinear model. We built the observation matrix Y by stacking the warped face images. Each column of Y has the warped face image of a specific subject with all illuminations and each row has the warped face image of all the subject with a specific illumination. We take $S = 6$, $C = 15$, and $K = 6237$ for Eq. (10), where K is the dimension of the warped face image.

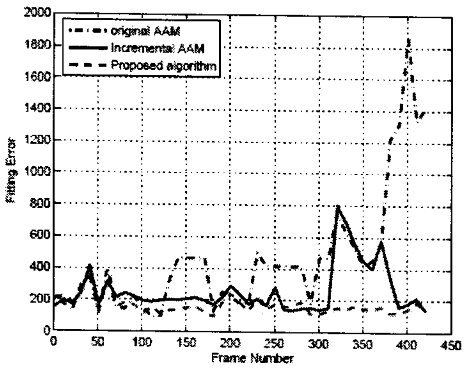

Fig. 3. Fitting error of each algorithm

5.2 Comparison of fitting performance

We compared the fitting performance of the original AAM, incremental AAM and proposed method. For evaluating the fitting performance of each algorithm, we recorded a image sequence which has varying illumination. It has gradual illumination change at the middle of the sequence, then sudden illumination change starts near frame index 300. Fig. 3. shows the fitting error of each algorithm. The fitting error is measured as the sum of squared error between the fitted shape point and the ground-truth shape point:

$$Error = \sum_{i=1}^{N} \sqrt{(x_i^{fit} - x_i^{g})^2 + (y_i^{fit} - y_i^{g})^2}, \qquad (17)$$

where (x_i^{fit}, y_i^{fit}) is the i-th fitted shape point, (x_i^{g}, y_i^{g}) is the i-th ground-truth shape point, and N is the number of shape points[1]. From the figure, we can see that the fitting error of original AAM increases as the illumination changes. On the contrary, the fitting error of incremental AAM does not increase under the gradual change of illumination(from frame index 120 to 180 and from frame index 220 to 290), since the AOM adapts to the change of appearance variation of the face image and the update of appearance base is take place. However, when the illumination changes drastically after the frame index 300, the AOM cannot adapt to the change and as a result the fitting error increases like the original AAM. In case of the proposed algorithm, the fitting error is not changed much through the entire image sequence.

Fig. 4. shows the fitting result of each algorithm when the illumination changes drastically. We can see that the original AAM cannot fit the model to the input face image since the trained appearance bases do not contain illumination variations. The incremental AAM also cannot fit the model to the input face since the AOM cannot adapt to the rapid change of the illumination. However, the fitting result of the proposed algorithm is stable since we updated the appearance bases using the synthesized illumination face images at the first frame.

There were 32 updates of appearance bases for incremental AAM, while only 7 updates (1 for input image and 6 for synthesized illumination images) were taken place for the proposed algorithm. Moreover, the incremental AAM should determine the goodness of the fitting result and update the AOM parameters for every frame. Thus, our proposed algorithm has less computation time than the incremental AAM.

6. Conclusion

In this paper, we proposed a novel scheme to update the appearance bases. First, for a new image, the AAM is initialized using a face and eye detector and fitted to the input image. Next, the algorithm determines whether the fitted face is a newly appeared person. If the fitted face is a newly appeared person, the illuminated face images are synthesized and the appearance bases of AAM is updated using these images. Then, the updated AAM is used for the next input image. By doing this, the appearance bases can fit to the input image accurately, even when the illumination condition changes drastically. The experimental results show that our proposed algorithm improves the fitting performance over both the incremental AAM and original AAM.

Acknowledgements

This work was partially supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University. Also, this work was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy (MOCIE).

References

[1] T. Cootes, G. Edwards and C. Taylor, "Active appearance models." In Proc. of European Conference on Computer

[1] In this experiment, we used 70 shape points.

Vision, pp. 484-498, 1998.

[2] S. Lee, J. Sung and D. Kim, "Incremental update of linear appearance models and its application to aam: incremental aam," In Proc. of ICIAR 2007, 2007.

[3] S. Zhou, R. Chellappa and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive model in particle filter," IEEE Trans. on Image Processing, vol. 13, pp. 1491-1506, 2004.

[4] P. Hall, D. Marshall and R. Martin, "Incremental eigenanalysis for classification," In: British Machine Vision Conference, 1998.

[5] J. Tenenbaum and W. Freeman, "Separating style and content with bilinear models," Neural Computation vol. 12, pp. 1247-1283, 2000.

[6] T. Sim and T. Kanade, "Combining models and exemplars for face recognition: An illuminating example," In: Workshop on Models versus Exemplars in Computer Vision, 2001.

[7] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible illumination conditions," International Journal of Computer Vision vol. 28, pp. 245-260, 1998.

[8] A. Shashua and T. Riklin-Raviv, "The quotient image: class-based re-rendering and recognition with varying illuminations," IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 23, pp. 129-139, 2001.

Frame 320 Frame 360 Frame 372 Frame 390

(a) Fitting result of original AAM.

Frame 320 Frame 360 Frame 372 Frame 390

(b) Fitting result of incremental AAM.

Frame 320 Frame 360 Frame 372 Frame 390

(c) Fitting result of incremental AAM using synthesized illuminated images.

Fig. 4. Comparison of the fitting result of each algorithm.