Enhance the Alignment Accuracy of Active Shape Models Using Elastic Graph Matching

Sanqiang Zhao[1, 2], Wen Gao[2], Shiguang Shan[2], Baocai Yin[1]

[1)] Multimedia and Intelligent Software Technology Lab, Beijing University of Technology, Beijing, China, 100022

[2)] ICT-ISVISION JDL for Face Recognition, Institute of Computing Technology, CAS, Beijing, China, 100080

{sqzhao, wgao, sgshan}@jdl.ac.cn, ybc@bjut.edu.cn

Abstract

Active Shape Model (ASM) is one of the most popular methods for image alignment. To improve its matching accuracy, in this paper, ASM searching method is combined with a simplified Elastic Bunch Graph Matching (EBGM) algorithm. Considering that EBGM is too time-consuming, landmarks are grouped into contour points and inner points, and inner points are further separated into several groups according to the distribution around salient features. For contour points, the original local derivative profile matching is exploited. While for every group of inner points, two pre-defined control points are searched by EBGM, and then used to adjust other points in the same group by using an affine transformation. Experimental results have shown that the proposed method greatly improves the alignment accuracy of ASM with only a little increase of time requirement since EBGM is only applied to a few control points.

Keywords

Face Alignment, Active Shape Models (ASMs), Active Appearance Models (AAMs), Elastic Bunch Graph Matching (EBGM), Point Distribution Model (PDM)

1 Introduction

The accurate localization and alignment of facial feature points is of great importance for face recognition, animation and tracking, etc. Over the last two decades, various methods have been proposed to deal with this task. After Kass et al proposed Active Contour Models (ACMs) [1], or snakes in 1987, Cootes and Taylor's Active Shape Models (ASMs) [2] and later Active Appearance Models (AAMs) [3] have proven to be among the most powerful tools in this field.

ASM and AAM are both based on statistical models. In ASM, the derivative profile, or local texture along the normal of the shape boundary, is exploited to model the local feature of each landmark point, and used to search each landmark position. Then the global shape models are applied to adjust the local search result based on the statistical shape model. Yet, as further research shows [4, 5, 8], due to its ambiguous local-texture searching strategy, ASM performs most successfully only when the object or structure class is fairly consistent in shape and intensity appearance. On the other hand, AAM [3] combines global statistical shape models with texture constraints to build appearance models and uses a linear prediction to obtain the appearance parameters for optimization; thus the shape can be calculated by minimizing the texture reconstruction error. To some extent, AAM may give a quite good match to image texture, but when the target image and background vary significantly, it is still unable to locate feature landmarks accurately. Meanwhile, both the training process and the searching procedure of AAM are quite complex and slow.

Therefore, much improvement was put forward to advance ASM for face alignment. Ginneken et al [4] presented a non-linear gray-level appearance instead of the original derivative profile to model the local texture, and got a better matching result. Rogers et al [5] suggested a robust parameter estimation method using M-estimator and random sampling approaches to evaluate the shape parameters more reliably. However, ASM still depends heavily on the initialization and may easily be stuck in local minima.

Among other methods, Wiskott et al [6] constructed a stack-like structure, called Face Bunch Graph, and used it to search the whole image to find the pre-defined feature points. Through the iterating distortion of the graph, this Elastic Bunch Graph Matching (EBGM) algorithm can tolerate a certain degree of pose and expression changes

and demonstrate a successful result. However, since this method is Gabor feature based, its time-consuming nodes searching process in the entire image region to large extent confines its further progress.

On account of all above, in our work, the landmarks are sorted into contour points and inner points, and the inner points are further separated into several groups according to the distribution around salient features in a face region. For the contour points, the original normal-based searching strategy is operated; for every group of inner points, two control points are chosen, and a simplified EBGM algorithm is utilized to adjust them. Fully understanding that the EBGM algorithm is quite a time-consuming process, we select the control points very carefully and representatively. When the new positions of the control points are identified, an affine transformation is applied to adjust other points in every group. This process is seamlessly combined with the ASM iterating searching algorithm and thus the accuracy can be greatly improved. The experimental results show that this method performs significantly better, with a speed not much slower than standard ASM algorithm.

The remaining part of this paper is organized as follows. In Section 2, the fundamentals of standard ASM are described. The simplified EBGM algorithm is presented in Section 3. After the detailed demonstration of our algorithm in Section 4, the experimental results are listed in Section 5, and the last section concludes the paper.

2 Standard ASM

ASM is definitely one of the best-suited approaches in the shape detecting task. The advantage of ASM is that it allows for considerable variability but still specific to the class of objects or structure they intend to represent. The following is a brief description of the standard ASM technique.

Given a set of annotated images, the manually labeled n landmark points in each image can be represented as a vector

$$\mathbf{x} = [x_0, y_0, x_1, y_1, \cdots, x_{n-1}, y_{n-1}]^T. \quad (1)$$

After aligning these vectors into a common coordinate, principal component analysis is applied to get a set of orthogonal basis $\mathbf{P}$. Every aligned shape can then be approximately represented as

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{Pb}, \quad (2)$$

where $\mathbf{b}$ is the shape parameters vector. Meanwhile, in the standard ASM, Cootes et al [2] used a normalized derivative profile to build the local texture models for each landmark. The similarity between a normalized derivative search profile $\mathbf{g}_s$ and the model profile $\bar{\mathbf{g}}$ can be measured by its Mahalanobis distance, which is a weighted square error function

$$f(\mathbf{g}_s) = (\mathbf{g}_s - \bar{\mathbf{g}})^T \mathbf{C}_g^{-1} (\mathbf{g}_s - \bar{\mathbf{g}}), \quad (3)$$

where $\mathbf{C}_g$ is the covariance matrix of the normalized derivative profile.

Based on the Point Distribution Model (PDM) and the gray level model, the search progress can then be operated. After initialization, each landmark in the model is optimized by selecting the point with a minimum distance mentioned above in the direction perpendicular to the contour within a certain range. As the result of new shape is possibly implausible, PDM is used to adjust the shape parameters. Such procedure is repeated until no considerable change is observed.

3 A Simplified EBGM Algorithm

3.1 Gabor Jets Representation

Gabor wavelet consists of a planer sinusoid multiplied by a Gaussian envelope, and it is fundamental to the EBGM algorithm. Daugman [7] pioneered the 2-D Gabor wavelet representation in computer vision in 1980's. A family of complex-valued 2-D Gabor kernel functions can be expressed as

$$j_j\left(\vec{x}\right) = \frac{k_j^2}{s^2}\exp\left(-\frac{k_j^2 x^2}{2s^2}\right)\left[\exp\left(i\vec{k}_j\vec{x}\right) - \exp\left(-\frac{s^2}{2}\right)\right], \quad (4)$$

with

$$\vec{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos f_u \\ k_v \sin f_u \end{pmatrix}, k_v = 2^{-\frac{v+2}{2}p}, f_u = u\frac{p}{8}. \quad (5)$$

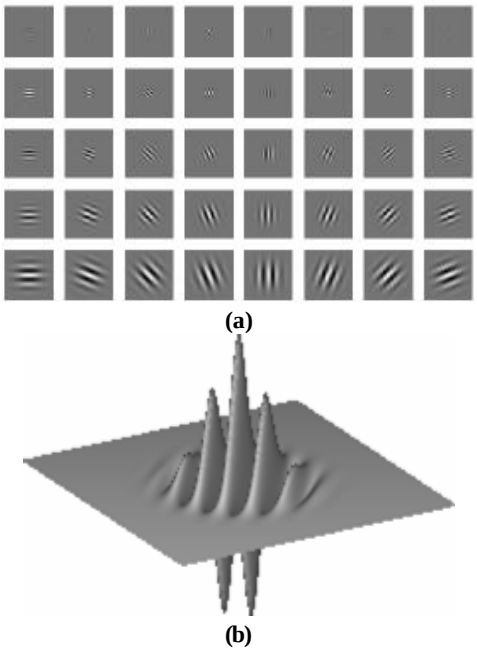


Figure 1. Gabor Wavelet. (a) Real part of 40 Gabor kernels. (b) The shape of a wavelet.

And we employ 5 frequencies $v=0,\cdots,4$ and 8 orientations $u=0,\cdots,7$ with index $j=u+8v$. Figure 1(a) shows the real part of these 40 Gabor kernels. Thus, we can use a *jet* to describe the local frequency information around each feature point. If given a pixel $\vec{x}$ in an image with gray level $I(\vec{x})$, a convolution process

$$J_j(\vec{x}) = \int I(\vec{x}')j_j(\vec{x}-\vec{x}')d^2\vec{x}' \quad (6)$$

will produce 40 complex coefficients. This collection of 40 Gabor coefficients obtained for every single image point is referred to as a *jet*, which is used to represent the local features.

Since complex, a jet can be written as

$$J_j = a_j \exp(if_j), \quad (7)$$

where magnitudes $a_j(\vec{x})$ vary slowly with position and phases $f_j(\vec{x})$ rotate at a rate approximately determined by the spatial frequency $\vec{k}_j$ of the kernels. When we obtain many jets referring to the same feature point in different images, this set of jets is called a *bunch*.

3.2 Displacement Estimation

To measure the similarity for two jets, two similarity functions are applied. The first one is magnitude similarity measure, which does not use the phase information:

$$S_a(J,J') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}. \quad (8)$$

It only computes the similarity of the energy of the frequencies and thus varies smoothly with the change of the position. As it is tolerant of small displacements, this method can be easily confused and may result in an incorrect spatial feature. The second function is phase similarity measure:

$$S_f(J,J') = \frac{\sum_j a_j a'_j \cos(f_j - f'_j - \vec{d}\vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}, \quad (9)$$

which is based on the similarity of both the magnitudes and the phase angles. This method uses a displacement $\vec{d}$ to compensate for small shifts from the original location, considering the phase information's rapid change with the displacement. The advantage of this displacement estimated similarity measure is that it can greatly reduce the number of Gabor wavelet convolutions that need to be computed by extracting one jet to represent the region near the landmark. To determine the displacement $\vec{d}$ between jets J and J', we can use a two-term Taylor expansion to approximate the cosine term in the phase similarity function (9) by

$$S_f(J,J') \approx \frac{\sum_j a_j a'_j [1-0.5(f_j - f'_j - \vec{d}\vec{k}_j)^2]}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}. \quad (10)$$

So it is possible to solve the displacement by maximizing (10). The final equation for calculating $\vec{d}$ is:

$$\vec{d}(J,J') = \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \frac{1}{\Gamma_{xx}\Gamma_{yy} - \Gamma_{xy}\Gamma_{yx}} \begin{pmatrix} \Gamma_{yy} & -\Gamma_{yx} \\ -\Gamma_{xy} & \Gamma_{xx} \end{pmatrix} \begin{pmatrix} \Phi_x \\ \Phi_y \end{pmatrix} \quad (11)$$

if $\Gamma_{xx}\Gamma_{yy} - \Gamma_{xy}\Gamma_{yx} \neq 0$, with

$$\Phi_x = \sum_j a_j a'_j k_{jx}(f_j - f'_j), \Gamma_{xy} = \sum_j a_j a'_j k_{jx} k_{jy}. \quad (12)$$

Equation (11) yields a straightforward method for estimating the displacement between two jets taken from object locations close enough that their Gabor kernels are highly overlapping. When all five frequency levels are eventually used, the displacement will be estimated accurately and quickly [6].

3.3 A Simplified EBGM Algorithm

As [6] presented, the standard EBGM algorithm uses a coarse-to-fine procedure to find the landmarks and thus to extract from the image a graph that can maximize the similarity with the Face Bunch Graph (FBG), which is a stack-like structure that serves as a representation of both jets and edges information. However, this procedure is quite a slow process, because the final optimal image graph is obtained by distorting FBG in various sizes and positions over the entire image region, and that actually needs every pre-defined feature point to search in a large range.

Fortunately, in most practical applications, ASM can provide a quite close approximation to the initial landmark positions, which makes the rough searching stage in the EBGM unnecessary. Besides, the PDM in ASM can successfully confine the landmark locations to a class of acceptable geometrical structures, so edge constraints in EBGM is not needed to participate in the similarity measure for our algorithm. In this sense, based on [6], a simplified EBGM algorithm is presented as follows:

1) Initialize the rough position of the landmarks by ASM.

2) For a landmark $\vec{x}$ that needs to be accurately positioned, calculate its jet J using equation (6).

3) From the bunch corresponding to that landmark, select the i^{th} jet as a model jet J_{mi}, and use equation (11) to calculate an estimated displacement $\vec{d}_i$ between jets J and J_{mi}, with all five frequency levels used.

4) Use equation (9) to compute the similarity S_{fi} between J and J_{mi} with the displacement $\vec{d}_i$

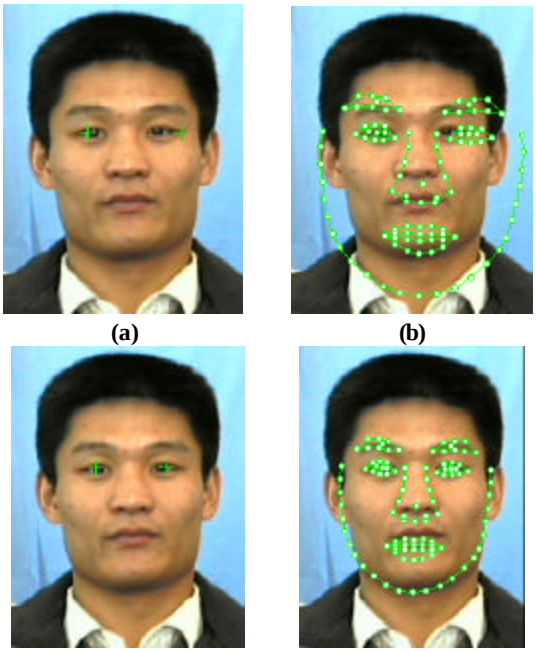
obtained from the former step. Then find another jet in the bunch and go to step 3) to repeat the above procedures until all the jets in the bunch have been operated.

5) Find the highest S_{fi} value as the best one: $i^* = \arg\max_i \{S_{fi}\}$. And the corresponding $\vec{d}_{i^*}$ is used as the optimal displacement for the landmark $\vec{x}$.

6) Change to the next landmark to be adjusted and go to step 2) for another searching process.

One point to be noticed is that, in step 4), we do not take a jet J' from position $\vec{x}' = \vec{x} + \vec{d}_i$; instead, we approximate the similarity between J' and J_{mi} by computing the similarity between J and J_{mi} with the displacement $\vec{d}_i$. This is because equation (9) can estimate the similarity as if J' was extracted from a displacement $\vec{d}_i$ from its current location [9]. As a whole, jet computation for a landmark searching process is only once in this simplified EBGM algorithm, which will reduce computational effort greatly.

4 ASM Combined with EBGM Searching Strategy



(a) **(b)** **(c)** **(d)**

Figure 2. Irises' slight inaccuracy will induce ASM's significant initialization error. (a) An example of irises' inaccuracy. (b) Significant ASM initialization error from (a). (c) EBGM refinement to (a). (d) ASM initialization from (c).

4.1 Further Adjustment for Iris Locations

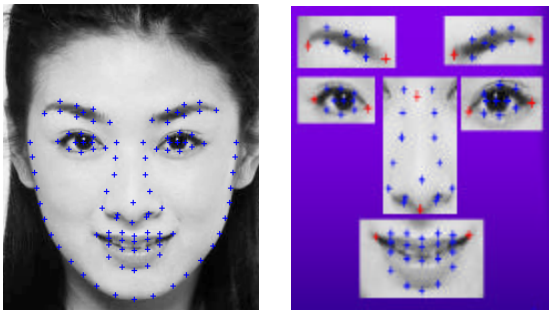
The proper initialization of the mean shape is critical to ASM searching progress in the sense that a good initialization would be less possible to lead to incorrect local minima. The standard ASM partially solves this problem by multi-resolution strategy at the cost of more expenditure spent during training in every resolution level.

In most current face detection systems, two irises, thanks to their natural distinct feature, can be located easily and quickly, though sometimes not very accurately (Figure 2a). And it has been proved that two irises can provide enough information to initialize the translation, the scale and the rotation angle for the mean shape model [8].

Experiments show irises' slight inaccuracy will induce ASM's significant initialization error (Figure 2b). Therefore, we use the simplified EBGM algorithm in Section 3.3 to refine the face detector's result. Figure 2(c) gives an example. This refinement result also reveals the accurate searching performance of this algorithm.

4.2 Feature-Based Landmark Grouping and Alignment

The purpose of this research is to use EBGM searching method to refine the ASM searching performance so that we can get a more desirable localization. So it is natural to think about changing all the normal-based searching strategy into EBGM algorithm, just as [10] proposed. Unfortunately, even when EBGM is simplified, it is still quite slower than ASM's normal-based searching method; moreover, experiments show that EBGM algorithm does not always perform better than the normal-based searching method in some outlying points along the image border [11].



(a) **(b)**

Figure 3. Face landmarks grouping. (a) A manually labeled face image with 103 landmarks. (b) 78 inner points separated into six groups; red crosses representing control points.

Under this condition, we sort the face landmarks (Figure 3a) into contour points and inner points; the inner points

are further separated into several groups based on the distribution around salient features in a face region. Figure 3(b) shows an example of six groups of inner points corresponding to two brow areas, two eye areas, one nose area and one mouth area. For the contour points, the original normal-based searching strategy is operated; as to every group of inner points, we select two control points, which are used to control all the other inner points in the same group. Since these two control points are very important, the simplified EBGM algorithm is utilized to refine them. To speed the searching performance, we define only 11 representative control points (Figure 3b) totally in all six groups. As an exception, one of the control points in the nose area is omitted, for it can be estimated from other control points in the two eye areas.

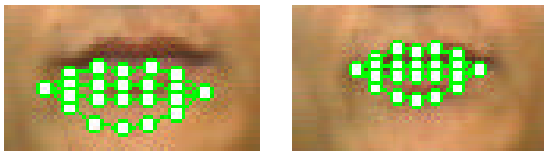
After the new positions of control points are identified by the simplified EBGM algorithm, we apply an affine transformation to adjust other points in every corresponding inner group. Actually this is a problem of alignment between two sets of feature points. We define this alignment as the rotation, scaling and translation which minimizes the sum of squared distances between pairs of corresponding points, then such a coordinate transformation in 2-D can be written in the following form:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} s\cos q & -s\sin q \\ s\sin q & s\cos q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = \begin{pmatrix} a & -b & t_x \\ b & a & t_y \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \quad (13)$$

Since we have two corresponding sets of points (two control points before and after refinement), the equation (13) can be further rewritten as follows:

$$\begin{pmatrix} x_1 & -y_1 & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ x_2 & -y_2 & 1 & 0 \\ y_2 & y_2 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ t_x \\ t_y \end{pmatrix} = \begin{pmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \end{pmatrix} \quad (14)$$

By equation (14), we can compute the corresponding affine transformation parameters, and then the remaining inner points' positions in the same group will be easily calculated by equation (13). Figure 4 gives an instance of refinement to landmarks in the mouth area using our algorithm.



(a) (b)

Figure 4. Landmark alignment. (a) Initial landmarks in the mouth area. (b) Result of EBGM adjustment to (a).

4.3 ASM Combined with EBGM Searching Strategy

In standard ASM matching process, every point along the normal of the shape contour is searched to find a best one, and the accuracy of every searching step is very important to the final result. Based on the above landmark grouping and alignment, our algorithm reaches a compromise between computational effort and searching accuracy.

The entire searching procedure of our algorithm is listed as follows:

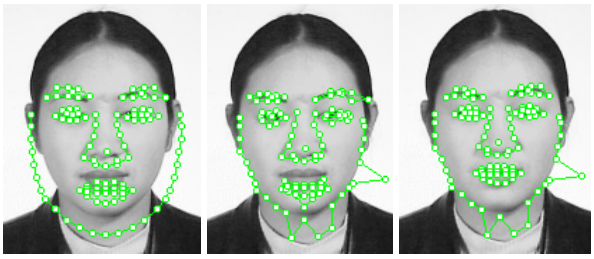
1) Initialize the starting shape by the model shape and two iris locations, which are identified by a face detector and refined by the simplified EBGM algorithm. See Figure 5(a).

2) For the contour points, use the standard normal-based searching method to find the required movements.

3) For every group of inner points, the simplified EBGM algorithm is operated for searching the two control points. After that, the method in Section 4.2 is used to adjust other points in the same group. We then get a new shape (Figure 5c).

4) Calculate additional pose and shape parameter changes required to move the model shape as close as possible to the new shape.

5) Update the shape and pose parameters by the above changes, and act on the model shape. Till now, we finish an iteration step. Go to step 2) for another matching circle until no considerable change is observed.



(a) (b) (c)

Figure 5. Comparison between ASM and our algorithm searching process. (a) Initialization. (b) Standard ASM searching result. (c) ASM combined with EBGM searching strategy.

5 Experimental Results

Our experiment is based on a 500 manually labeled faces database, most of which are near frontal. We have normalized 350 of them for training the PDM and others for testing; we also selected 70 representative face images for training the EBGM bunches. To evaluate the performance of our algorithm, the average Euclidean distance error is calculated by the following equation:

$$E = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n}\sum_{j=1}^{n}\sqrt{(x_{ij}-x'_{ij})^2+(y_{ij}-y'_{ij})^2}\right), \qquad (15)$$

where N is the total number of test images, and n is the number of the landmarks in one face image. (x_{ij}, y_{ij}) is the j^{th} manually labeled landmark location of the i^{th} test image; and (x'_{ij}, y'_{ij}) is the corresponding j^{th} landmark location we calculated.

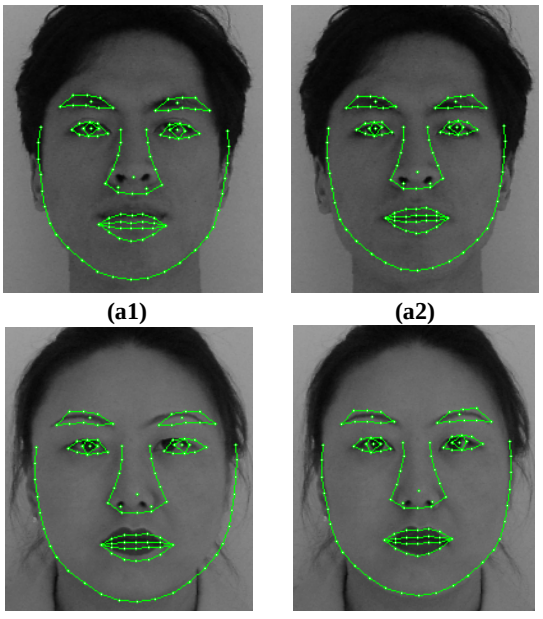
We also calculate the overall improvement percentage of our algorithm to the standard ASM by:

$$I = \frac{E_{ASM} - E_{ASM-EBGM}}{E_{ASM}} \times 100\% . \qquad (16)$$

Table 1 lists our final experimental results based on 150 test images. The average error for standard ASM is about 3.08 pixels. After we apply EBGM to refine two iris locations, the error is reduced to 2.54 pixels. The overall improvement of our algorithm is about 39.9% compared with the standard ASM.

Table 1. Performance of our algorithm

Method	Average Error	Improvement
ASM	3.08	
ASM with iris -refinement	2.54	17.5%
ASM combined with EBGM	1.85	39.9%



(a1) **(a2)** **(b1)** **(b2)**

Figure 6. Matching comparison between standard ASM and our algorithm. (a1), (b1) Result of ASM. (a2), (b2) Result of our algorithm.

Figure 6 lists some matching results of our algorithm as well as standard ASM. Through comparison between them, we can see our algorithm improves the searching accuracy significantly, especially that of the inner points.

6 Conclusions and Further Work

In this research, we combine the standard ASM algorithm with a simplified EBGM to improve the alignment accuracy. Because EBGM has the advantage of Gabor wavelet representation, the landmarks can be more accurately localized compared with using gray level representation. In order to reach a compromise between the searching efficiency and the alignment accuracy, we classify all the landmarks into contour points and inner points, and further separate the inner points into several groups according to their distribution around salient features in a face region. We apply the simplified EBGM algorithm to accurately localize the two control points of every group, and then, an affine transformation is used to adjust other points in the corresponding inner group. Experimental results show that our algorithm performs much better than the standard ASM.

Future work will be focused on perfecting a robust EBGM searching strategy to deal with face images with rotation in depth and poor quality. Meanwhile, applying the proposed method to face recognition is also a practical research effort.

Acknowledgements

This research is partially supported by National Hi-Tech Program of China (No. 2001AA114160, No. 2001AA114190 and No. 2002AA118010), Natural Science Foundation of China (No. 60375007 and No. 60332010), National 973 Program of China (No. 2001CCA03300), Natural Science Foundation of Beijing of China (No. D070601-01) and ISVISION Technologies Co., Ltd.

References

[1] M. Kass, A. Witkin and D. Terzopoulos, "Active Contour Models", 1st International Conference on Computer Vision, London, June, 1987, pp. 259-268.

[2] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application", Computer Vision and Image Understanding, 61(1), 1995, pp. 38-59.

[3] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," Proceeding of the 5th European Conference on Computer Vision, vol. 2, 1998, pp. 484-498.

[4] B.V. Ginneken, A.F. Frangi et al, "A Non-linear Gray-level Appearance Model Improves Active Shape

Model Segmentation", IEEE Workshop on Mathematical Models in Biomedical Image Analysis, MMBIA 2001, pp. 205-212.

[5] M. Rogers and J. Graham, "Robust Active Shape Model Search", Proceedings of the European Conference on Computer Vision. May, 2002.

[6] L. Wiskott, J.M. Fellous, N. Kruger et al, "Face Recognition by Elastic Graph Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, July, 1997.

[7] J.G., Daugman, "Complete Discrete 2-D Gabor Transform by Neural Network for Image Analysis and Compression", IEEE Transactions on Acoustics, Speech and Signal Processing, 36(7), 1988, pp. 1169-1179.

[8] W. Wang, S. Shan, W. Gao and B. Cao. "An Improved Active Shape Model For Face Alignment", The 4th International Conference on Multi-modal Interface, IEEE ICMI 2002, Pittsburgh, USA, pp. 523-528, Oct., 2002.

[9] D.S. Bolme, "Elastic Bunch Graph Matching", Masters Thesis, CSU Computer Science Department, June, 2003.

[10] F. Jiao, S. Li, H. Y. Shum and D. Schuurmans, "Face Alignment Using Statistical Models and Wavelet Features", International Conference on Computer Vision and Pattern Recognition, CVPR 2003.

[11] B. Zhang, W. Gao, S. Shan and W. Wang. "Constraint Shape Model Using Edge Constraint And Gabor Wavelet Based Search", 4th International Conference on Audio and Video Based Biometric Person Authentication, AVBPA 2003.