Extended Fitting Methods of Active Shape Model for the Location of Facial Feature Points

Chunhua Du[1], Jie Yang[1], Qiang Wu[2], Tianhao Zhang[1], Huahua Wang[1], Lu Chen[1], and Zheng Wu[1]

[1] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China
[2] Department of Computer Systems, University of Technology, Sydney, 2007, Australia

Abstract. In this study, we propose three extended fitting methods to the standard ASM(active shape model). Firstly, profiles are extended from 1D to 2D; Secondly, profiles of different landmarks are constructed individually; Thirdly, length of the profiles is determined adaptively with the change of level during searching, and the displacements in the last level are constrained. Each method and the combination of three methods are tested on the SJTU(Shanghai Jiaotong University) face database. In all cases, compared to the standard ASM, each method improves the accuracy or speed in a way, and the combination of three methods improves the accuracy and speed greatly.

1 Introduction

The location of facial landmarks plays a very important role in face research. Most facial features location methods can be broadly divided into two categories: One is local methods[1],[2],[3]. The other is global methods[4],[5],[6]. Deformable templates, active shape model(ASM) and active appearance model(AAM) all belong to the global methods. Compared to local methods, global methods are more robust, the most important thing is that the number of facial landmarks detected by global methods is greatly more than that of the local methods. ASM has been successfully applied in many areas[7],[8]. However, its drawbacks in accuracy and speed limit its further application. First, standard ASM searches the new position in 1D profile. However, the true new position may not lie in 1D profile, it therefore may bring to some error. Secondly, local structure of each landmark in standard ASM is generated identically. However, all landmarks are characterized by different features. Construct structures alike is unreasonable. Thirdly, length of profiles in different levels equals to each other during searching. In fact, the displacements become to decrease with the increase of the level. Hence length of the profiles in different levels doesn't need to be equal.

This paper presents three methods to address the above problems. Firstly, the 1D profile is extended from 1D to 2D; Secondly, local structures for the different landmarks are constructed according to their characteristic features; Thirdly, we decrease the length of the profiles when it comes to the next level during the fitting. We also constrain the displacements in the last level.

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 610 – 618, 2006.

The rest of the paper is organized as follows. Section 2 introduces the standard ASM. Three extended fitting methods are described in Section 3. Section 4 gives the comparison between the proposed methods and the standard ASM, experimental results are also presented. Section 5 is the conclusion of the paper.

2 Active Shape Models

ASM[4]is a statistic model, which includes the construction and fitting. The parameters used in this paper are listed in Table 1.

Table 1. Parameters of ASM(stand ASM and our methods). Values given between parentheses are used in our experiments.

k	Number of landmarks(60)
n	Number of training images(200)
s	Number of images in the testing set(2069)
t	Number of modes in the shape model(21)
m	Number of points in sub-profile on either side of the landmark (4)
l	Number of points in profile either side of the landmark for searching, length of the profile is 2l+1(11 for the first, second method and single-resolution standard ASM. 11,9 and 7 for different levels of the third method and the combined method)
L	Number of levels(3)
N	Number of iterations(50 for single-resolution scheme,5 for each level of multi-resolution scheme)
T	Threshold of the displacement of the last level for the third method and the combined method(2)

2.1 Construction of the ASM

For each face image in face database, mark k facial landmark $(x_1, y_1),...,(x_k, y_k)$ manually on picture as is shown in Fig. 1. Each landmark with the same index number

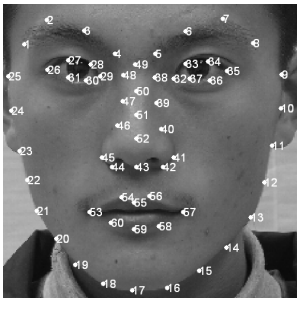

Fig. 1. Facial landmarks

in different images stands for the same feature, and all the landmarks of an image give a shape vector

$$x = (x_1, y_1, x_2, y_2, ..., x_k, y_k) \tag{1}$$

To make the shape model independent of the size, position and orientation, all shape vectors are aligned[4]. Principal component analysis(PCA) is then applied to the aligned shape vectors and we can get the mean shape

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

and the covariance

$$S = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^T (x_i - \bar{x}) \tag{3}$$

Then the eigenvalues $(\lambda_1, \lambda_2, ..., \lambda_s)$ and the eigenvector $(p_1, p_2, ..., p_s)$ are computed. Select the first *t* eigenvectors and eigenvalues to satisfy the inequation:

$$\sum_{i=1}^{t} \lambda_i / \sum_{i=1}^{s} \lambda_i \geq 0.95 V_T \tag{4}$$

where V_T is the sum of all the eigenvalues. Then a shape vector can be approximated by

$$x \approx \bar{x} + Pb \tag{5}$$

where b is a vector of weights, computed by

$$b = P^T (x - \bar{x}) \tag{6}$$

In order to find the new positions for facial landmarks, local structure needs to be constructed in advance. As for a landmark (x_i, y_i), sample m pixels on either side then give a profile. Then local structure can be built by using the normalized first derivatives of this profile. Denoting the normalized derivative profiles $g_1, g_2, ..., g_n$, the mean profile $\bar{g}$, covariance matrix S_g, we then can compute the Mahalanobis distance between a new profile g_i and the mean profile

$$f(g_i) = (g_i - \bar{g}) S_g^{-1} (g_i - \bar{g})^T \tag{7}$$

2.2 ASM Fitting

ASM fitting can be realized by two iterative steps:(1)finding a new position for each landmark;(2)updating *b* and the parameters of the affine transformation.

At first, sample *l* pixels either side of the landmark and give a profile alike as before, select a sub-profile and compute the Mahalanobis distance. The center of the sub-profile with the minimal Mahalanobis distance is the new position. Find the new

position for each landmark and compute the displacement for each landmark to get a displacement vector $dX = (dX_1, dX_2, ..., dX_k)$ .Then the following equation is used to update the affine transformation parameters and *b*:

$$X = M(s,\theta)[x] + X_c \tag{8}$$

According to equation (8), we can get

$$M(s(1+ds),(\theta+d\theta))[x+dx] + (X_c + dX_c) = (X + dX) \tag{9}$$

and

$$M(s(1+ds),(\theta+d\theta))[x+dx] = M(s,\theta)[x] + dX + X_c - (X_c + dX_c) \tag{10}$$

According to equation (5), db is utilized such that $x + dx = \bar{x} + P(b + db)$ and $db = P^{-1}dx$. Then affine transformation parameters and *b* are updated as follows:

$$X_c = X_c + w_t dX_c, Y_c = Y_c + w_t dY_c, \theta = \theta + w_\theta d\theta, s = s(1+w_s ds), b = b + W_b db \tag{11}$$

where w_t, w_θ, w_s, W_b are scalar weights.

3 Extended Fitting Methods of the Active Shape Models

Although there are many ways to refine the standard ASM, we discover that three methods can best improve the accuracy and speed of it.

3.1 Extend the Profile from 1D to 2D

It is well known that there is a key idea in the standard ASM: finding a new position for each landmark. This finding accuracy affects the whole accuracy of ASM directly. In Fig. 2, the target position of the current landmark is point P1, whereas the best position that can be located is point P2. To tackle this problem, we propose to extend profile from1D to 2D as shown in Fig. 3.

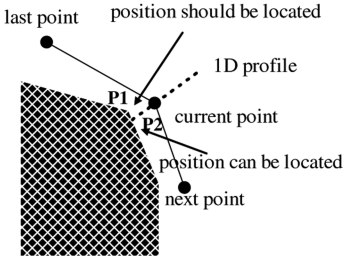

Fig. 2. 1D profile

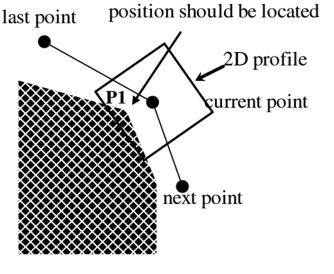

Fig. 3. 2D profile

Since the profile has been extended to 2D, local structure should be constructed in another way: for each landmark, select a square with side length of m and center of current landmark, such that there are n squares, then compute the average gray-scale of each pixel within the square, and get a 2D profile g_o. A square with the side length of l and the center of the current landmark is selected during the fitting. We define a function f_o to replace the Mahalanobis distance:

$$f_o = \frac{1}{m \times m} \sum_{i=1}^{m} \sum_{j=1}^{m} abs(g^{ij}{}_o - temp_o^{ij}) \tag{12}$$

where g_o^{ij} is the gray-scale of pixel lies in the *ith* row , *jth* column of 2D profile of *oth* landmark, $temp_o^{ij}$ is counterpart of the sub-profile of the *oth* landmark, f_o represents the similarity between the 2D profile and the 2D sub-profile. The center of the sub-profile is viewed as the new position for the current landmark if f_o reaches the minimum.

3.2 Construct the Local Structures for Landmarks According to Their Feature

In standard ASM, local structures of all landmarks are constructed identically. However, local structures of landmarks don't express their true feature. Accordingly, the accuracy of location will subject from this kind of local structure. In order to improve the accuracy, we propose to construct the local structures for landmarks individually. For the landmarks around the irises and eyebrows, we construct the local structures not only with profiles, but also with gray-scale information. And a new function is used

$$f(g_i) = gray \times (g_i - \overline{g}) S_g^{-1} (g_i - \overline{g})^T \tag{13}$$

where $gray$ is the gray-scale of the landmarks. Then for the landmarks on the contour of chin and mouth, we construct their local structures with profiles and edge information, and use another function

$$f(g_i) = -edge \times (g_i - \overline{g}) S_g^{-1} (g_i - \overline{g})^T \tag{14}$$

where $edge$ is the magnitude of the landmark of the gradient image.

3.3 Adjust the Length of the Profile Adaptively

In the standard ASM, length of the profile in different levels is fixed. Experiments indicate that the displacements of the landmarks in the first level are very large, and the displacements begin to decrease with the increase of the level. Thus we propose to reduce the length of the profile with the increase of the level and constrain the displacements in the last level by setting a threshold T.

4 Comparison and Experiments

To compare the proposed methods with the standard ASM, experiments are conducted on the SJTU face database, which includes 2269 labeled frontal face images and the size of image is 640×480. We divide the SJTU face database into training set and testing set. The former and the latter contain 200 and 2069 images respectively. All experiments are conducted on a P4 2.8GHz machine with Matlab implementations. In order to compare our methods with the standard ASM quantitatively, we first define two functions, average error

$$E_{ave} = \frac{1}{s} \times \frac{1}{k} \sum_{i=1}^{n} \sum_{j=1}^{k} dis(X(i,j) - pos(i,j)) \tag{15}$$

and average computation time

$$ave\_time = \frac{1}{s} \sum_{i=1}^{s} time_i \tag{16}$$

where $X(i,j)$ is the manually marked position of the *jth* landmark of the *ith* image in the testing set, $pos(i,j)$ is the position located by different methods, $dis(X(i,j) - pos(i,j))$ is the Euclidean Distance between $X(i,j)$ and $pos(i,j)$, $time_i$ is the computation time for the fitting of the *ith* face.

At first, we compare the first method with the standard ASM. All other aspects of the two algorithms are identical except the shape of the profiles. Both use 50 iterations for fitting. Since profiles have been extended from 1D to 2D, the corresponding searching spaces are also extended. Therefore, the probability of finding the new positions will be increased and the accuracy of whole fitting will be increased accordingly. We plot the average error against the number of iteration N= 5, 10,…, 50 in Fig. 4, which shows that the average error of the first method is lower than that of the standard ASM during the fitting. Furthermore, it needs less iteration to converge. From Fig. 4 we can also see that the first method has converged after 25 iterations, and the counterpart of the standard ASM is 35. However, on the other hand, the average computation time for two methods is 0.30s and 0.33s. It is not strange because the profile has been extended from 1D to 2D, and the corresponding searching space also be extended form 1D to 2D, it will come down to more pixels and need more computations.

Then for the second method, we construct the local structure for landmark individually. Local structure of the second method has more information. Thus, the new local structure can discriminate the landmarks more efficiently. We plot the average error against the number of the iterations in Fig. 5. It shows that the second method also needs fewer iterations to converge. On the other hand, the second needs more computation since the functions of equation (14) and (15) need a more multiplication operation. The average computation time for the second method and the standard ASM is 0.31s and 0.30s.

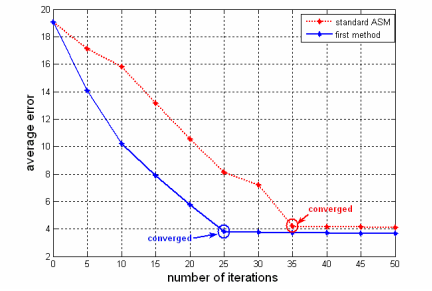

Fig. 4. Average error against the number of iterations

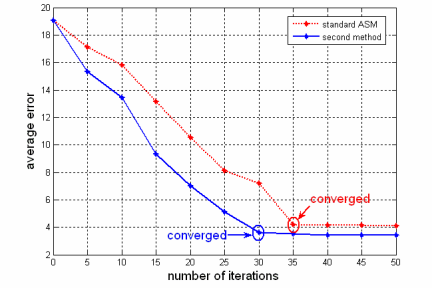

Fig. 5. Average error against the number of iterations

For the third method, the number of pixels each side of the landmark in the first, second and last level is 11, 9 and 7. The displacement threshold T in the last level is 2. For the standard ASM, the number of pixels each side of the current landmark in all levels is 11. For both methods, the number of levels and the iterations of each level is 3 and 5 in turn. We plot the average error against the iterations in Fig. 6, from which we can see that the standard ASM has converged after two levels(1-10 iterations). Fig. 6 also indicates that the average error of the two methods is the same in the first level. When it comes to the second level, we shorten the profiles. As for the third method, the number of iterations for converge is only 8, which is less than that of the standard ASM. The average computation time for the third method and standard ASM is 0.11s and 0.08s

At last, we combine all three methods to give a combined method. Since the profiles are extended to 2D and the local structures of the landmarks are constructed differently, the accuracy of finding the new positions is enhanced, which leads to a great increase in the accuracy of the whole fitting. As the length of the second and third level in the combined method is shortened, the speed for the fitting is reduced. The average computation time of the combined ASM and standard ASM is 0.09s and 0.11s. We also plot the average error against the iterations Fig. 7. It is clear that the combined method needs 7 iterations to fitting.

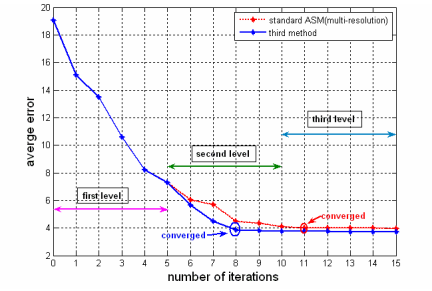

Fig. 6. Average error against the number of iterations

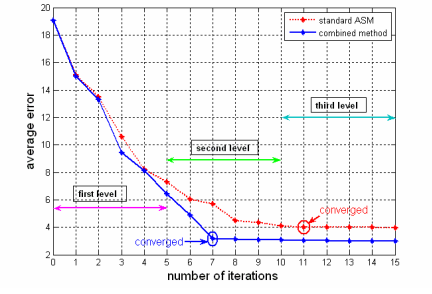

Fig. 7. Average error against the number of iterations

5 Conclusions

In this paper we have presented three methods to the fitting of the standard ASM to locate the facial landmarks. Experimental results demonstrate that each of method can improve the accuracy or speed in a way, and the combined method improves the accuracy and speed greatly, when compared with the standard ASM.

References

1. Zhi-Hua Zhou and Xin Geng, Projection functions for eye detection, Pattern Recognition, vol. 37, no. 5, pp. 1049-1056, 2004.
2. Kawaguchi T and Rizon M, Iris detection using intensity and edge information, Pattern Recognition, vol. 36, no. 2, pp. 549-562, 2003.
3. Shu-Hung Leung, Shi-Lin Wang, and Wing-Hong Lau, Lip image segmentation using fuzzy clustering incorporating an elliptic shape function, IEEE Transactions on Image Processing, vol. 13, no. 1, pp. 51-62, 2004.
4. T.F.Cootes, C.J.Taylor, D.H.Cooper, and J.Graham, Active Shape Models - Their Training and Application, Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59, 1995.

5. Alan L.Yuille, Peter W.Hallinan, and David S.Cohen, Feature extraction from faces using deformable templates, International Journal of Computer Vision, vol. 8, no. 2, pp. 99-111, 1992.
6. T.F.Cootes, G. J. Edwards, and C. J. Taylor, Active Appearance Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, 2001.
7. Christos Davatzikos, Xiaodong Tao, and Dinggang Shen, Hierarchical active shape models, using the wavelet transform, IEEE Transactions on Medical Imaging, vol. 22, no. 3, pp. 414-423, 2003.
8. B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, Active shape model segmentation with optimal features, IEEE Transactions on Medical Imaging, vol. 21, no. 8, pp. 924-933, 2002.