An Improved Active Shape Model for Face Alignment

Wei Wang[*1], Shiguang Shan[2], Wen Gao[2,3], Bo Cao[2], Baocai Yin[1]

College of Computer Science, Beijing Polytechnic University, Beijing, China, 100022

ICT-YCNC FRJDL, Institute of Computing Technology, CAS, Beijing, China, 100080

School of computer science, Harbin Institute of Technology, Harbin, China, 150001

{wwang, sgshan, wgao, bcao}@jdl.ac.cn, yinbc@bjpu.edu.cn

Abstract

In this paper, we present several improvements on the conventional Active Shape Models (ASM) for face alignment. Despite the accuracy and robustness of the ASMs in the image alignment, its performance depends heavily on the initial parameters of the shape model, as well as the local texture model for each landmark and the corresponding local matching strategy. In this work, to improve the ASMs for face alignment, several measures are taken. First, salient facial features, such as the eyes and the mouth, are localized based on a face detector. These salient features are then utilized to initialize the shape model and provide region constraints on the subsequent iterative shape searching. Secondly, we exploit the edge information to construct better local texture models for the landmarks on the face contour. The edge intensity at the contour landmark is used as a self-adaptive weight when calculating the Mahalanobis distance between the candidate profile and the reference one. Thirdly, to avoid their unreasonable shift from the pre-localized salient features, landmarks around the salient features are adjusted before applying the global subspace constraints. Experiments on a database containing 300 labeled face images show that the proposed method performs significantly better than traditional ASMs.

Keywords: Face Recognition, Face Alignment, Active Shape Models

1. Introduction

Face recognition has a variety of potential applications in multi-modal interface, commerce and law enforcement, such as mug-shot database matching, identity authentication, access control, information security, and surveillance. Related research activities have significantly increased over the past few years [1-4]. Though much progress has been made in the past few years and several successful commercial face recognition systems have emerged in the Biometric market, the FRVT2000 shows that practical face recognition system is still a great challenge [4]. One of the bottlenecks in a practical face recognition system is the face alignment problem, i.e., labeling some (even each) pixel with the high-level semantics related to its facial physiologic configuration. For instance, the pixel located at the eye corner should be labeled with "eye corner". Face alignment is also vital to gaze tracking, pose estimation, expression classification etc.

A variety of approaches have been proposed to deal with the face alignment problem including the active contour [5], deformable template [6], Active Shape Models [7,8], Face Bunch Graph Matching [9] and Active Appearance Models [10]. Active shape models, proposed by Cootes and Taylor [7, 8], have been shown to be an effective tool to understand the configuration of the face images, as well as the medical images. Some methods have been proposed to improve the ASM. Bram van Ginneken *et al*.[11] proposed to use a non-linear gray-level appearance instead of the first derivative profile to model the local texture in order to get a better matching result. Mike Rogers *et al*.[12] proposed a robust parameter estimation method using M-estimator and random sampling approaches to estimate shape parameter more accurately. However, ASMs still depend heavily on the initial model and local texture modeling.

In our face alignment task, we localize some salient facial features including the eyes and the mouth by using a features localization method based on a face detector. These salient features are then utilized to initialize the shape model and provide region constraints on the subsequent iterative shape searching. First, the two eyes are used for a better initial shape with "known" scale, translation and rotation parameters. Since the initial shape is near to the target with this initialization strategy, no multi-resolution method is needed again, and the search

[*] This work is completed when the author is as a visiting student at the ICT-YCNC FRJDL, Institute of Computing Technology, CAS.

progress can still converge in less iterations. Thus the efficiency of the algorithm can be greatly improved. Second, we use these pre-localized landmarks to constrain the match result. Landmarks around the salient features are adjusted before applying the global subspace constraints to avoid their possible unreasonable shift, caused by the unreliable local texture matching.

We also exploit the edge information to construct better local texture models for the landmarks on the face contour, based on the observation that the landmarks on the face contour usually locate at a strong edge. So, it is more probable that the target point should have bigger edge intensity. To make use of this point, in this paper, the edge intensity at the contour landmark is used for deriving a self-adaptive weight when calculating the Mahalanobis distance between the candidate profile and the reference one.

To compare the proposed improvement with the standard ASM, experiments are conducted on a database containing 300 labeled face images, which show that the proposed method performs better.

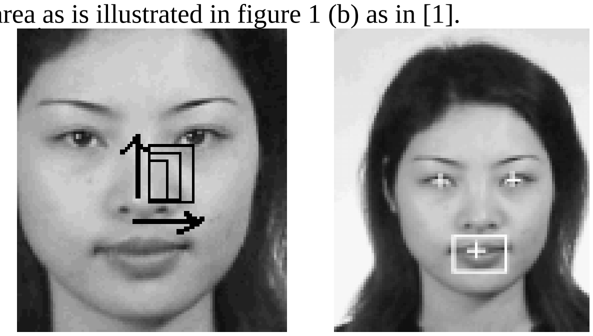
The remaining part of the paper is organized as follows: In Section 2, the method for iris and mouth localization is presented. Section 3 described in detail the proposed improvements on ASM. The experiments are presented in Section 4. A short conclusion as to the proposed method is drawn in the last section followed by some discussion and future works.

2. Salient facial features localization

Of all the facial features, the two irises are the most distinct organs. Therefore, they are relatively easier to locate. And they can provide enough information to initialize the translation, the scale and the in-image-plane rotation parameters for the shape model. Mouth center can also be detected in an expected area determined by iris locations to provide further constraints on the active shape modeling procedure.

In our system, the face detector is derived from a rule-based method in [13]. Then in the detected face region, an iris location approach is adopted. The basic idea of the iris locator is based on the observations that, in a face image, the gray level intensity in the iris area tends to be lower than that of other parts. The main idea of the iris locating strategy is to use this information to search iris in the expected area provided by face detection method. From the estimated nose center of the detected face, the method searches the possible iris edge by gradually growing a rectangle (as figure 1 (a) illustrated). And then localize the iris center in an expected region by searching the pixel that has minimal averaging intensity. By this strategy, an accurate position of iris can be detected for face images under variations in pose, facial expression and lighting condition. Refer to [14] for detailed algorithm.

After we have located the irises, we can estimate the rough position of the mouth center. First the center of the mouth should be on the line perpendicular to iris line. Secondly, the gray level of the region should be smaller than other part. So we can locate the center of the mouth by analyzing the integral projection curve in the expected area as is illustrated in figure 1 (b) as in [1].



(a) (b)

Figure 1. Salient features localization. (a) Search iris by gradually growing a rectangle. (b) The detected iris and mouth center, where the rectangle is the integral region.

3. Improved active shape model

Based on the located three salient facial features, in this section, we improve the standard active shape model by introducing region constraints. Also, the edge constraints are introduced to improve the local texture model for the landmarks on the face contour.

3.1. Shape model initialization according to the iris locations

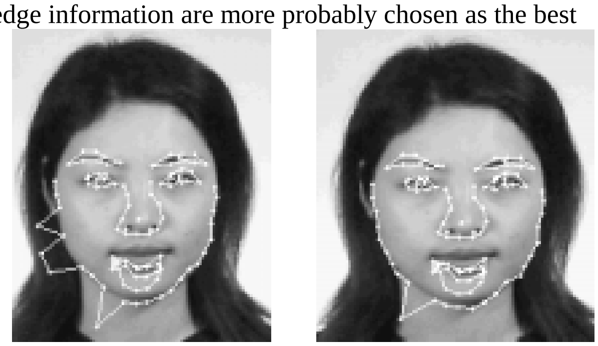
In ASM search progress, the initialization of the mean shape is very important, because a good initialization would less lead to incorrect local minimal. With the positions of the two irises, we can calculate the parameters of the scale, rotation, and translation for the target face in the image. In ASM, the shape of a model is described as $X = (x_1, y_{1,} \cdots x_n, y_n)$, where $x_i, y_i (1 \le i \le n)$ is the coordinate of the i-th landmark. The initial value of the model points X_i in the image can be calculated by $X_i = T_{x_t, y_t, s, \theta}(\overline{X} + \Phi b)$, where the function $T_{x_t, y_t, s, \theta}$ is a transform including rotating by θ, scaling by s , and a translation of (x_t, y_t) . The initial value of statistical shape parameter b can be set as zero. As figure 5 shows, the initial shape is very close to its real values. So multi-resolution strategy is not needed, and better results can be achieved in less convergence times.

3.2. Edge constraint in local texture model matching

In ASM, local texture model matching is conducted under the assumption that the normalized first derivative profile, $\{g_i\}$, satisfies a Gaussian distribution. The matching degree of a probe sample, g_s, to the reference model is given by $f(g_s) = (g_s - \overline{g})^T S_g^{-1}(g_s - \overline{g})$ where $\overline{g}$ is the mean of $\{g_i\}(1 \le i \le N)$, and N is the number of images for establishing Point Distribution Model (PDM), S_g is the covariance. The main idea of finding the matching point is to minimize $f(g_s)$ which is equivalent to minimizing the probability that g_s comes from the distribution. There is no conclusion that this should be an optimal choice. For most images interpretation task, the landmarks are usually selected in the points that have strong edge information. So the destination points should have also strong edge information. These points tend to lie on the edges extracted by an edge operator. Thus we can adjust the matching method by adding a weight to the malahanobis distance function:

$$f(g_s) = (c-k)(g_s - \overline{g})^T S_g^{-1}(g_s - \overline{g}),$$

where c is a constant, k is the Sobel edge intensity at the target point. With this improvement, points with strong edge information are more probably chosen as the best candidate. For face images, the landmarks on the contour part have clearly strong edge intensity, while the landmarks on other parts are not necessary with this property. Therefore, we only apply the edge constraints to the landmarks on the face contour. As illustrated in figure 2, most mismatched points on the contour in the standard ASM have been correctly matched by applying our edge constraints on the local texture model matching.



(a) (b)

Figure 2. Comparison of the local texture matching with and without edge constraints. (a) Matching result of the standard ASM. (b) Matching result with edge constraint

3.3. Adjustment on local search window

The size of local search window is also an important part in the searching strategy. If the search region is too small, more probably the target point may be missed in the search progress; on the other hand, if it is too large, some unwanted point may be included, which will lead to bad result. In ASM, the problem is partly solved by multi-resolution strategy. As we already have the location of irises, the overall search length can be determined according to the distance between the two irises. We choose a longer searching distance when the distance between tow iris points is long. Further more, the window size should also be different for the landmarks on different part of the shape. In a face image, because the model points in the constrained parts (eyes and mouth) are more probably to be close to their real position, smaller window size is enough. However, in other parts, larger window would be more appropriate in order to guarantee the true point within the search region.

3.4. Re-adjusting landmarks around salient key features

In ASM matching progress, each point is searched along the normal of the shape contour for a best matched point. After matching we can calculate the difference between X (shape before matching) and X' (shape after matching) $dX = X' - X$. With dX, We can calculate the shape parameter adjustment db and the affine transformation parameters to generate a new shape for next iteration. The accuracy of the new shape X' is very important to the final result, so we should make the new shape X' as accurate as possible. Since we have already have the accurate position of the irises and the mouth center point, we can fully make use of these prior information to adjust X'. As figure 3 illustrated, the iris point P_0 should approximately locate in the center point of gravity of the whole eye shape calculated by $P_i(1 \le i \le 10)$, $P_0 x \approx 1/10\sum_{i=1}^{10} P_i x$, $P_0 y \approx 1/10\sum_{i=1}^{10} P_i y$. In ASM matching progress some landmark points may shift away from its true position due to the unreliable local search, which may lead to a bad result after a few iterations. So it is necessary for us to make the match result as accurate as possible. As described above, the gravity center of the eye landmarks should approximately locate in the prior detected iris point. Thus we can adjust the matched eye points to make their gravity center equal to the previously detected iris point. $P_i^{'}x = P_i x - dx,\ P_i^{'}y = P_i y - dy \quad (1 \le i \le 10)$, where $P_i x$ and $P_i y$ are the coordinate of the landmarks around eye after matching progress, $P_i^{'}x$ and $P^{'}_{i}y$ are their coordinate after adjustment, dx and dy are the difference between the detected iris and the gravity center of the eye landmarks after matching progress.

The method can prevent the landmarks around eyes from shifting away too distant from their actual locations. Similar strategy can be applied to the landmarks in the mouth region. As it is illustrated in figure 4, more robust and more accurate result can be achieved when such constraints applied.

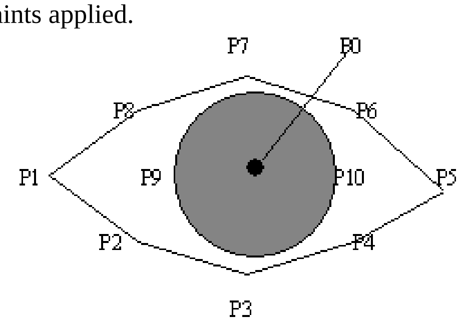
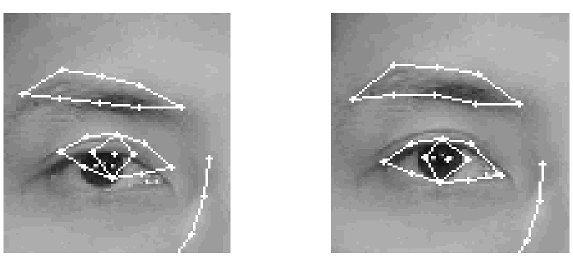


Figure 3. The iris center point is approximately the gravity center calculated from the eye landmarks, where *P0* is the detected iris center, *P1-P10* is the landmark points around eye.



(a) (b)

Figure 4 Effect of salient feature constraints (a) Bad search result without eye position constraint. (b) Result with eye constraint.

4. Experiments

In this part we introduce our experimental system, and show the experimental results. We have also compared our improved ASM with standard ASM based on a deliberate performance evaluation method.

4.1. Face labeling tool and database

In order to reduce the error brought by manual labeling, a tool is designed for landmark points labeling. In the labeling progress, some clearly pre-defined key feature points, such as the eye corners, mouth corners, etc., are first labeled, based on which, then, the x-coordinates or y-coordinates of some other points are constrained. We argue that the Point Distribution Model established from these training images can preserve more useful statistical information by filtering out the possible factitious labeling errors.

Using this tool, we have labeled 300 near frontal face images to establish the PDM and evaluate the proposed methods. We chose the images base on the following principles. First, the depth rotation of face and facial expression shouldn't be very significant to ensure the detection of face and iris location will not fail. Second, the difference of the actual face size in image shouldn't be too big. The size of most of the images we choose is 240×320 pixels. Among these images, there are 117 females and 183 males. Most of them are oriental.

4.2. Performance evaluation

Performance evaluation is really an important problem for different approaches to face alignment. In this paper, we propose to evaluate the performance by using average error, which is defined as the following distance between the manually labeled shape and the resulting shape of our ASM:

$$E = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n}\sum_{j=1}^{n} dist(P_{ij}, P_{ij}^{'})\right)$$

where N is the total number of the probe images, n is the number of the landmark points in the shape (for our case, n=103), P_{ij} is the j-th landmark point in the manually labeled shape of the i-th test image manually labeled, $P_{ij}^{'}$ is the j-th landmark point in the resulting shape of ASM for the i-th test image . The function $dist(p1, p2)$ is the Euclidean distance between the two points.

4.3. Experimental results

To evaluate our method, experiments are conducted on the 300 faces database abovementioned. In order to evaluate the performance more accurately and sufficiently, the leave-one-out strategy is adopted. For each probe image, the remaining 299 images in the 300 faces database are utilized to establish the PDM. Then, the PDM is utilized by the proposed ASM to search a shape for the one not participating in establishing the PDM.

As Table 1 shows, the average error for standard ASM is about 2.14 pixels. After applying edge constraint on face counter, the average error is reduced to 2.11 pixels. The error is reduced to 2.06 pixels after applying constraint on eyes and mouth. With the above two schemes applied the error is reduced to 2.03 pixels per point. The overall improvement of our strategies to the stand ASM by our performance system is about5.14%.

Some results of our improved ASM are presented in this part. Figure5 demonstrates the initial location of the mean shape and the search result.

Table 1. Performance comparison of different methods

Method	Average Error	Improvement
Standard ASM	2.14	-----
Edge constraint	2.11	1.40%
Region constraint	2.06	3.73%
Two schemes applied	2.03	5.14%

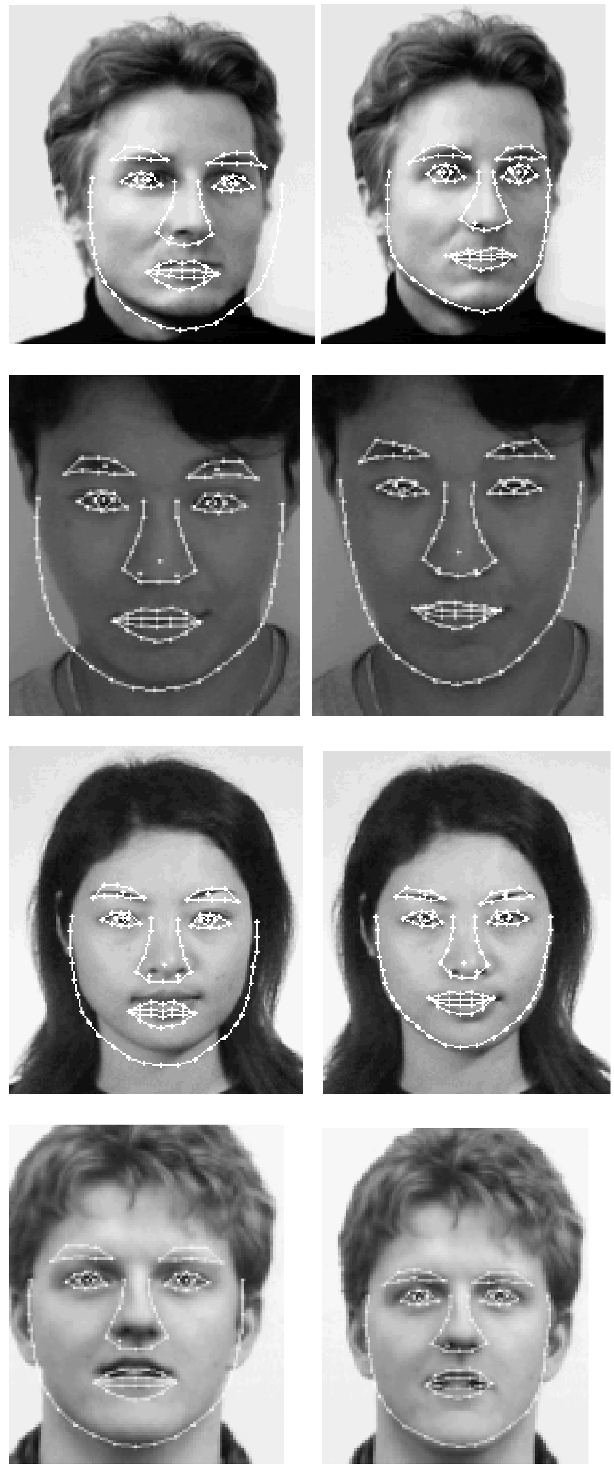
5. Conclusions and future work

In this paper, to solve the face alignment problem, we propose some improvements on the conventional Active Shape Model (ASM). The main contributions of the paper are:

1. Salient facial features, such as the eyes and the mouth, are localized and utilized to initialize the shape model.
2. The edge information is exploited to construct better local texture models for the landmarks on the face contour. The edge intensity at the contour landmark is used as a self-adaptive weight when calculating the Mahalanobis distance between the candidate profile and the reference one.
3. To avoid their unreasonable shift from the pre-localized salient features, landmarks around the salient features are adjusted before applying the global subspace constraints.

The proposed approaches are compared with the standard ASMs, which show that the proposed method performs better.

Our future work will be focused on the accuracy of the salient features localization and some more deliberate local texture models for the landmarks such as eye corners, mouth corners. Global texture constraints are also to be exploited. How to use the result of ASM for face recognition is also a future research effort.



(a) (b)

Figure5. Result of the improved ASM. (a) Initialized average shape model by located iris. (b) Final results of the proposed method.

Acknowledgement

This research is sponsored partly by Natural Science Foundation of China (No.69789301), National Hi-Tech Program of China (No.2001AA114190), SiChuan Chengdu YinChen Net. Co. (YCNC), 100 Talents Foundation of Chinese Academy of Sciences and Natural Science Foundation of Beijing (4011001).

Reference

[1] R.Brunelli and T.Poggio, “Face Recognition: Features versus Template”, *TPAMI*, 15(10), pp1042-1052, 1993

[2] R.Chellappa, C.L.Wilson ect. “Human and Machine Recognition of faces: A survey”, *Proc. of the IEEE*, 83(5), pp705-740, 1995.5

[3] P.J.Phillips, H.Moon, etc. “The FERET Evaluation Methodology for Face-Recognition Algorithms”, *IEEE TPAMI*, Vol.22, No.10, pp1090-1104, 2000

[4] D.M.Blackburn, M. Bone, P.J. Phillips, Facial Recognition Vendor Test 2000: evaluation Report, Feb.2001, htt://www.frvt.org/frvt2000/

[5] M.Kass, A.Witkin,and D.Terzopoulos. Snakes: Active contour models. Int. Journal of Computer Vision, pp.321-331, 1988

[6] A.L.Yuille, Deformable templates for face detection, J. Cogn. neurosci. 3, 59-70, 1991.

[7] T.F.Cootes, C.J.Taylor, D.Cooper, and J. Graham. Active shape models--their training and application. *Computer vision and image understanding*, 61(1): pp38-59, 1995.

[8] A.Lanitis, C.J.Taylor, T.F.Cootes. Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.19, no.7, pp.743-56, July 1997

[9] L. Wiskott, J.M. Fellous, N. Kruger, C.v.d.Malsburg, Face Recogniton by Elastic Bunch Graph Matching, IEEE Trans. On PAMI, Vol.19, No. 7, pp775-779, 1997

[10]T.F. Cootes, G.J. Edwards, C.J. Taylor, Active Appearance Models, Proc. European Conf. Computer Vision, vol. 2, pp. 484-498, 1998.

[11]B.V.Ginneken, A.F. Frangi etc, "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.

[12] M. Rogers and J. Graham. Robust Active Shape Model Search, to appear in the proceeding of *ECCV 2002*

[13]J.Miao, B.C.Yin, K.Q.Wang, et al, A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template, Pattern Recognition, 32(7), 1999.

[14]B.Cao, S.G.Shan, W.Gao, Localizing the iris center by region growing search, Proceeding of the *ICME2002*, (To appear)