

Cascade MR-ASM for Locating Facial Feature Points

Sicong Zhang, Lifang Wu*, and Ying Wang

School of Electronic Information and control Engineering
Beijing University of Technology, Beijing, China, 100022
Tel.: 8610-67396151,
{apatroit,lfwu}@bjut.edu.cn,

Abstract. Accurate and robust location of feature point is a difficult and challenging issue in face recognition. In this paper we propose a new approach of using a cascade of Multi-Resolution Active Shape Models (C-MR-ASM) to locate facial feature points. In our approach, more than one MR-ASMs are obtained from different subsets of training set automatically, and these MR-ASMs are integrated in a cascade to locate facial feature points. Experimental results show that our algorithm is more accurate than traditional MR-ASM. The contribution of this paper includes: 1, unlike traditional MR-ASM, the training set is divided into several subsets automatically based on the principle a trained model should describe all the samples in training set accurately. 2, we propose the new cascade framework, which integrates all the subset MR-ASM.

1 Introduction

Accurate and robust location of feature point is a difficult and *challenging* issue in face recognition. Cootes et al. proposed a Multi-Resolution Active Shape Models (MR-ASM) algorithm [1], which could extract specified shape efficiently and accurately.

Because of the robustness of MR-ASM, a lot of improved algorithms have been proposed. MR-ASM is a statistic model of shape variation, and so the accuracy of estimated distribution of training samples will determine the location result. Traditional improvements of MR-ASM focus on estimating and simulating the probability distribution [2,3,4]. In fact, due to the complicated variation of training samples, it is almost impossible that the training set conform to certain clustering property. It is therefore difficult to estimate and simulate the distribution of training samples accurately [6].

In this paper, we propose an algorithm that can easily avoid this problem. Firstly, the training set is divided into several subsets automatically based on the principle that a trained model should describe all the samples in training set accurately. Then each subset is used to train a subset MR-ASM. At last, all subset MR-ASMs can represent all the samples in training set more accurately than one single model.. To locate facial

* Corresponding author.

feature points, all the trained subset MR-ASMs are organized in cascade, so that a face image can be represented using the most suitable model.

Our algorithm comprises two stages: training of subset MR-ASMs and location of facial feature points using cascade MR-ASMs. We first describe our procedure to obtain a subset MR-ASM at the training stage. Firstly, we use the whole training set to train a model. We then use this model to fit the training samples: we pick a subset of training samples that can be accurately represented using the trained model. This subset of training samples is then used to train a new model. The above procedure is repeated until a model is obtained that can accurately enough (at a certain precision) represent all the samples in the subset under consideration. This model is model 1 and the corresponding subset is subset 1. The whole training set excluding subset1 is used as the new training set. By repeating the above procedure, model 2 and subset 2, model 3 and subset 3,...can be obtained in sequence. When the number of samples in all the subsets is more than 97.5% of the whole training set, about 2.5% reminder training samples are trained to get the final MR-ASM (MR-ASM 1) .

To locate feature points, all trained models are organized in a cascade. For an input image, the first model is used to fit it. If the texture of matched LMPs (LandMark Points) doesn't satisfy texture distribution of the training set, the second model is used, The procedure will continue until the texture of LMPs matched by a model satisfy the distribution of training set, the location of feature points will be obtained by the output of the model. If there is not a model (from MR-ASM 1 to MR-ASM l-1) by which the texture of output LMPs satisfy the distribution of training set, the face image will be fitted using the final model (MR-ASM l), and its output will be the location of feature points.

The following parts of this paper are organized as follows: In section 2, the training of MR-ASMs is described. In Section 3, the location of facial feature points using cascade MR-ASMs is described in details. Then experimental results are presented in section 4. Finally, the paper is concluded in section 5.

2 Training of MR-ASMs

In our work, 98 feature points in a face image are used as landmark points (LMPs), as shown in Fig. 1.

2.1 Training Texture at Each LMP

Training texture is to extract the gray level feature in the neighbours of each LMP. The neighbour of a LMP is a rectangle area, whose long side is parallel to the normal of the LMP, as shown in Fig. 2.

The gray level of pixels inside the neighbour of the i^{th} LMP is extracted one by one at the same order from inner face to outside, which forms a vector of gray level g_i . Two statistics mean g_i and covariance matrix Σ_i of vector of gray level g_i of all the training samples are computed respectively.

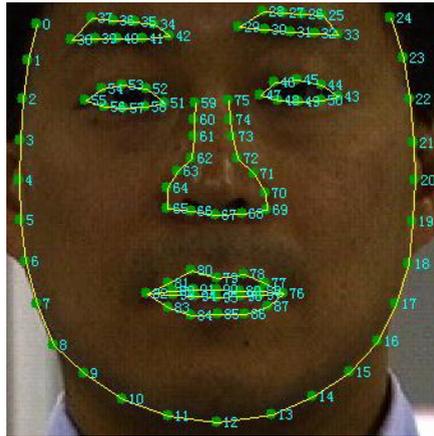


Fig. 1. Shape model in our work

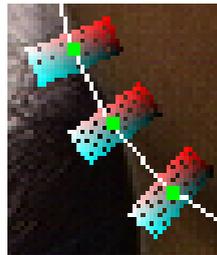


Fig. 2. Neighbour of LMPs

2.2 Training a MR-ASM

Training a MR-ASM comprises five steps [1]: 1. Labeling the training set (as Showing in Fig.2). 2. Extracting gray profile at every landmark point. 3. Aligning training set ASM. 4. Computing the statistics of aligned training set by PCA. 5. Repeat from step 1 to step 4 at each resolution level.

After these five steps, any shape X_i can be expressed as:

$$X_i = \bar{X} + P \cdot b \tag{1}$$

Where, \bar{X} is the mean ASM of training set, eigenvectors P represents the most important t modes of variation, b is the vector of coefficients.

2.3 Training Subset MR-ASMs

With the traditional methods estimating and simulating the probability distribution of training set were the focus. One would expect to find one or a set of perfect description

to represent the whole distribution. However in fact, it is very difficult to estimate and simulate the distribution of the whole training set [6].

Cootes et al [7] proposed a different way to solve this problem. They sorted training set into several subsets manually by the angle of the head. Samples of each subset are used to train a MR-ASM. The collection of these trained MR-ASMs as a whole represent the faces from a wide range of angles. We push Cootes’s idea further in this work.

With traditional methods one model is obtained from the training set. In fact, however, due to the complicated variation of samples in training set, it is difficult to represent all training samples accurately in the training set using one model. The naïve explanation for this phenomenon is that it is difficult to observe all of the things clearly from one perspective. On the contrary if we observe different subsets of things from different angles, it is possible that all the things can be observed clearly overall. Cootes’s idea is based on the above naïve understanding, where the subsets were created manually [7]. However Cootes’ approach is somewhat subjective and it is difficult to justify the subsets produced. In our work, the subsets are obtained automatically, based on the principle that a trained model should describe all the training samples accurately.

The framework of our algorithm is shown as follows:

Suppose, the whole training set is N_0 , the number of samples in N_0 is $Number_{N_0}$. The sequence of subset MR-ASMs is represented as $L_i, i = 0, 1, 2, \dots, k - 1$. Let $i = 0$.

1. Suppose that there are two subsets $SP = N_i, SN = \{\Phi\}$, From the training set SP , we can get a trained model $M_{i,0}$ by the steps in Section 2.2, and let $j = 0$

2. Using model $M_{i,j}$ to locate the LMPs of all the samples in subset SP . Computing the displacement from labeled LMPs for every samples and the average displacement $\overline{E_{i,j}}$ in subset SP .

3. If $\overline{E_{i,j}} < C_{th}$,

It means that the model $M_{i,j}$ can represent all the samples in subset SP , go to step5.

Otherwise,

Remove all the samples in which the average displacement of LMPs is bigger than $\overline{E_{i,j}}$ from subset SP , and insert them into subset SN .

4. Let $j = j + 1$, using the training set SP , we can train a model $M_{i,j}$, go to step 2.
5. do the following operation:

$$L_i = M_{i,j}, i = i + 1, N_i = SN$$

Count the number $Number_{SN}$ of samples in SN .

$$\text{If } Number_{SN} \leq 0.025 Number_{N_0}$$

It means that at least 97.5% training samples have been used,, the loop should be stopped. go to step 6.

Otherwise go to step 1.

6. From subset SN , we can train a model L_{k-1} .
7. Ordering the sequence of models $L_i, i = 0, 1, 2, \dots, k - 1$ by the number of samples in the corresponding subset SP , so that models at the front can represent more common variation in training set.

3 Location of the Facial Feature Points

In our approach the sequence of subset MR-ASMs are organized in a cascade, as shown in Fig. 3.

The fitting result is tested using the square of Mahalanobis distance:

$$M_{dist} = \sum_{i=1}^n (g_i - \overline{g_i})^T \Sigma_i^{-1} (g_i - \overline{g_i}) \tag{2}$$

Where, g_i is the vector of gray level at the i th LMP, $\overline{g_i}$ is the trained mean of gray level at the corresponding LMP. Σ_i^{-1} is the inverse of covariance matrix of the corresponding LMP.

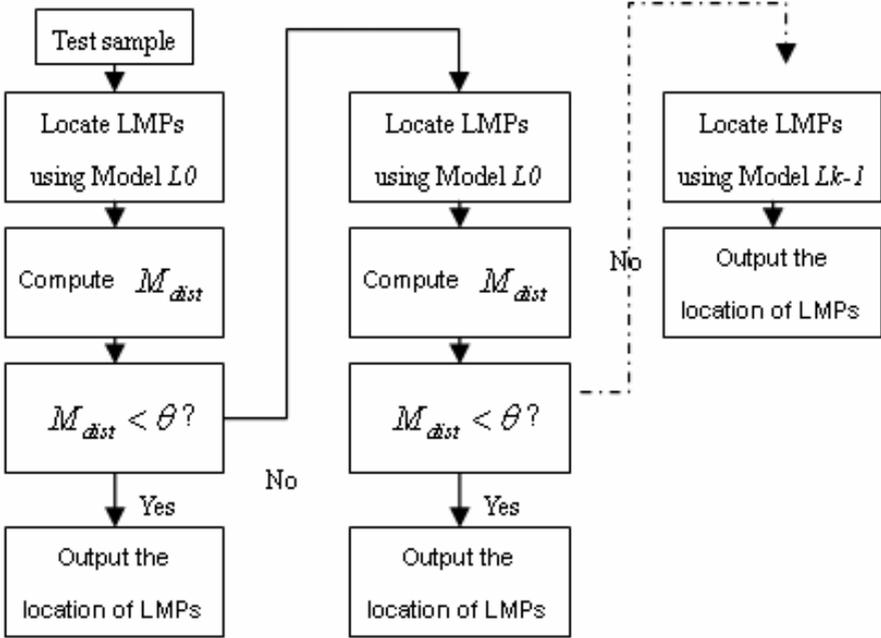


Fig. 3. The framework of feature point location

4 Experimental Results

The CMU-PIE face database is used to test the accuracy of our algorithm. A total of 153 face images of 17 people (each person uses 3 different images under 3 expressions) are chosen as training set. 36 unseen face images from the same 17 people are used as the test set. The size of images is 640x480 pixels. The average distance between two eyes is 90 pixels.

The PCA thresholds are set to 95% for every MR-ASM locators. Three gray levels are used in the MR-ASM algorithm. The length of gray profile at each LMP is set to 11 pixels. The number of searching positions is three in the both normal directions of face contour.

In the experiment, we trained five locators to compose the cascade location system. Details are shown in table 1.

Table 1. The cascade locator training information

Locator No.	Training sample selected	Location error for whole face	Location error for F.S.O.	Location error for chin
1	32	3.53	3.04	4.93
2	26	4.02	3.36	5.96
3	31	8.32	5.98	15.1
4	24	5.86	5.25	7.65
5	40	10.52	7.65	18.89

Where F.S.O. means five sense organs. Location error is measured in pixel.

Table 2 shows the first two shape parameters of each locator. It presents the main shape variation trend of each automatically selected subset of the train set.

Table 2. The shape parameters of the cascade locators

Locator No.	-3std	0	+3std	-3std	0	+3std
1						
2						
3						
4						
5						

The second to the fourth columns correspond to the first parameter. The fifth to the seventh columns correspond to the second parameter. It can be seen that different locators represent different variations of face shape.

We first compared the performance of testing the whole train set between our algorithm and traditional MR-ASM. The results are shown in table 3.

Table 3. The results compared with traditional MR-ASM for train set

	Face (pxl.)	F.S.O. (pxl.)	Cheek contour (pxl.)
MR-ASM	9.3	6.6	16.6
Our algorithm	4.98	4.00	7.32
Improvement	46.5%	39.4%	55.9%

It can be seen that, for all testing subjects, our algorithm is much more accurate than traditional MR-ASM. Further more, because of the less influence of background, the location results of eye, mouth, nose and eyebrow for both algorithms are the best. For the same reason, as to the feature points of cheek contour, both results are the worst.

Table 4. The results compared with traditional MR-ASM for testing set

	Face (pxl.)	F.S.O. (pxl.)	Cheek contour (pxl.)
MR-ASM	6.42	5.15	10.12
Our algorithm	5.53	4.65	8.11
Improvement	13.9%	9.7%	19.9%

Secondly we use both algorithms to test the unseen face image set. The results are shown in table 4.

5 Conclusion

With the development in the face recognition technology, face recognition in complex backgrounds becomes an interesting research topic, where accurate location of facial feature points is important.

In this paper an improved MR-ASM algorithm has been proposed. At the training stage, the training set is divided into several subset automatically based on the principle that a trained model should describe all training samples accurately, and more than one MR-ASM are trained to represent all training samples accurately. Then all the trained MR-ASM locators are integrated in cascade, so that an image can be identified by the most suitable MR-ASM locator.

Acknowledgement. This paper is supported by a Science and Technology project of the Beijing Municipal Education Commission under Grant No. KM200610005011 and an Academic Human Resources Development project of the Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR-IHLB).

References

1. Cootes, T.F., et al.: Multi-Resolution Search with Active Shape Models. In: Pattern Recognition 1994 Conference A: Computer Vision & Image Processing Proceedings of the 12th IAPR International Conference, 9-13 October 1994, vol. 1, pp. 610–612 (1994)
2. Hu, C., Chang, Y., Feris, R., Turk, M.: Manifold Based Analysis of Facial Expression. In: Computer Vision and Pattern Recognition Workshop, 2004 Conference, 27-02 June 2004, p. 81 (2004)
3. Rogers, M., Graham, J.: Robust Active Shape Model Search for Medical Image Analysis. Medical Image Understanding and Analysis, 81–84 (August 2002)
4. Romdhani, S., Psarrou, A., Gong, S.: Learning A Single Active Face Shape Model across Views, Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. In: Proceedings. International Workshop, 26-27 September 1999, pp. 31–38 (1999)
5. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference, vol. 1, pp. 511–518 (2001)

6. Cootes, T.F., Taylor, C., Wolfson, J.: Statistical Models of Appearance for Computer Vision. Image Analysis Unit, Imaging Science and Biomedical Engineering, University of Manchester. <http://www.medicine.manchester.ac.uk/isbe/>
7. Cootes, T., Walker, K., Taylor, C.: View-based active appearance model. In: Proceedings, International Conference on Automatic Face and Gesture Recognition. Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference, 28-30 March 2000, pp. 227–232. IEEE Computer Society Press, Los Alamitos (2000)