MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan

4-3

Robust Active Shape Model using AdaBoosted Histogram Classifiers

Yuanzhong Li
Imaging Software Technology Center
FUJI PHOTO FILM CO., LTD.
li_yuanzhong@fujifilm.co.jp

Wataru Ito
Imaging Software Technology Center
FUJI PHOTO FILM CO., LTD.
wataru_ito@fujifilm.co.jp

Abstract

Active Shape Model (ASM) has been shown to be a powerful tool to aid the interpretation of images, especially in face alignment. ASM local appearance model parameter estimation is based on the assumption that residuals between model fit and data have a Gaussian distribution. However, in face alignment, because of changes in illumination, different facial expressions and obstacles like mustaches and glasses, this assumption may be inaccurate. AdaBoost is widely used in face detection as a robust classification method, which does not need the Gaussian distribution assumption. In this paper, we model local appearances by using AdaBoosted histogram classifiers to solve the robustness problems, which have previously been encountered. Experimental results demonstrate the robustness of our method to align and locate facial features.

1 Introduction

Statistical shape models have been shown to be a powerful tool to aid the interpretation of images, especially in face alignment. Models represent the shape and variation of faces and can be used to impose priori constraints on face alignment. A frequently used formulation is the Active Shape Model (ASM) [1]. Many researches on applying ASM to face alignment have been done, but the technology still suffers from changes in illumination, different facial expressions and obstacles like mustaches and glasses.

To fit a model to data, parameters must be estimated in an optimal manner. Standard ASM parameter estimation minimizes the sum of squares of residuals between the model and the data. It has been widely recognized that least squares minimization only yields optimal results under the assumption of Gaussian distribution residuals. Under real conditions, a Gaussian model of residual distribution is seldom accurate. Face images taken in different conditions containing widely varying appearances and confusing local structures potentially give rise to non-Gaussian residuals. In [2], wavelet features and EM algorithm are used to model local appearances, but the assumption of Gaussian distribution is still used.

AdaBoost is widely used in face detection as a robust classification method [3] [4] [5][6]. AdaBoost does not need the Gaussian distribution assumption, and can be applied to non-linear classification problems by training weak classifiers. In this paper, to improve robustness of calculating landmark displacement, we use AdaBoosted histogram classifiers as local appearance models of each landmark. Experimental results show that compared to ASM, robustness of feature point displacement was improved greatly to changes in illumination, different facial expressions, and obstacles like mustaches and glasses. Additionally, our method is also robust to occlusions on faces, which may cause search failures in previous researches.

2 Active Shape Model

2.1 Statistical Shape Model

The ASM technique relies upon each object or image structure being represented by a set of points. Given a set of training images for a given object, points are manually placed in the same location on the object in each image. 103 landmarks used in this paper are shown in Fig. 1. The image is from CMU PIE database [7].

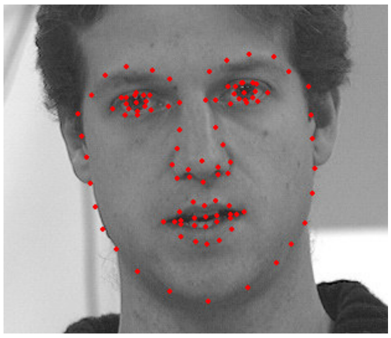

Fig. 1 Labeled image with 103 landmarks

The points from each image are represented as a vector and aligned to a common co-ordinate frame. Principle Component Analysis is applied to the aligned shape vector

$$S = \bar{S} + Pb \qquad (1)$$

where $\bar{S}$ is the mean shape vector, P is a set of principle components of shape variation and b is a vector of shape parameters.

The vector b defines a set of parameters for a deformable model. By varying the elements of b we can vary the shape using formulation (1). By applying bounds to the value of parameter b, we ensure that the generated shapes are similar to those in the novel training set.

The ASM search procedure is an iteration procedure. On each iteration, it uses the local appearance model to find a new shape and then updates the model parameters to best fit the new search shape [1].

2.2 Local Appearance Models

The local appearance models, which describe local image features around each landmark, are modeled as the first derivative of the sample profiles perpendicular to the landmark contour [1].

It is assumed that the local models are distributed as a Gaussian. For the jth landmark, we can derive the mean profile $\bar{g}_j$ and the sample covariance matrix S_j from the jth profile examples directly. The quality of fitting a feature vector g_s at test image location s to the jth model is given by calculating the Mahalanobis distance from the feature vector to the jth model mean.

$$f_j(g_s) = \left(g_s - \bar{g}_j\right)^t S_j^{-1}\left(g_s - \bar{g}_j\right) \tag{2}$$

Using local appearance models leads to fast convergence to the local image evidence. However, due to the variation of the illumination and obstacles, a feature point often cannot be accurately located. As a consequence, ASM tends to get stuck at local minima.

3 Model Local Appearance using AdaBoosted Histogram Classifiers

The most important thing in ASM is how to calculate landmark displacement. This calculation is based on intensity profiles, which are perpendicular to each landmark contour. Because of different illumination and obstacles, it is unreasonable to model them by using a Gaussian distribution model. We use AdaBoosted histogram classifiers to model the local appearances of each landmark by the following steps. The number of points selected from profile to make a histogram classifier and quantization value are determined empirically. To each landmark,

① Derive intensity profiles just at the manually labeled landmark position from training images as positive training samples.

② Derive intensity profiles at positions apart from the manually labeled landmark position from training images as negative training samples.

③ To each profile, calculate its intensity variance and quantize the intensities of the profile to 5 levels based on the intensity variance.

④ Make histogram classifiers based on quantized value combinations at 3 different positions of quantized profiles. Number of bins of each histogram is 5^3.

⑤ Train theses histogram classifiers based on training samples by using AdaBoost.

In step ⑤, let training samples be $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, where x_i is the training sample and y_i is the label for the sample (+1 for positive samples and –1 for negative samples). The AdaBoost algorithm trains weak histogram classifiers $f_m(x)$ so that the sum $F_M = \sum_{m=1}^{M} f_m(x)$ will have high classification accuracy. To archive this, we have to optimize the following objective function:

$$f_M(x) = \arg\min_f J\left(F_{M-1}(x) + f(x)\right) \tag{4}$$

It can be shown that the minimizer is

$$f_M(x) = \frac{1}{2} \ln \frac{P\left(y = +1 \mid x, w^{M-1}\right)}{P\left(y = -1 \mid x, w^{M-1}\right)} \tag{5}$$

where w^{M-1} are the weights given at time M and updated by using the following function:

$$w^M = w^{M-1} e^{-y f_M(x)} \tag{6}$$

The final function is $F(x) = \operatorname{sgn}\left(\sum_{m=1}^{M} f_m(x)\right)$, which should be positive for positive samples and negative for negative samples.

4 Search using AdaBoosted Histogram Classifiers

Except calculating landmark displacement, our search process is the same as ASM. Length of intensity profile at test image is 3 times long as the profiles used in training. To determine the displacements of each landmark, scores of each position in test profile are calculated through the following steps.

① To each position, apply the same quantization in training.

② To each position, calculate its score by using the AdaBoosted histogram classifiers based on the quantized intensities.

Locations with higher scores show higher confidence that the landmark should be displaced to. Therefore, we select the location with the highest score as the displacement location. One important thing is that we set displacement to 0 when the highest score is lower than 0. The reason is that a negative score means that the profile is a negative one. This is very effective to improve robustness to occlusions on faces, which have not been discussed by previous researches.

Three intensity profile, Mahalanobis distance and AdaBoost score distribution examples are shown in Fig. 2. In ASM, the position with the smallest Mahalanobis distance is selected as the displacement position. In our method, the position with the highest score is selected. Compared with Mahalanobis distance, the AdaBoost scores are much more accurate and reliable.

5 Experimental Results

We manually labeled 500 frontal face images, 300 as training images and 200 as test images. Distance between two eyes is almost 60 pixels. These faces include those taken in different illumination, with different expressions, without or with moustache and glasses. Our evaluation includes the following steps.

① Displace the mean face shape on each test image from the true position between -6 and +6 pixels randomly. Then, scale up or down the mean shape between 0.9 and 1.1 times randomly. Finally, rotate the shape between –5 and +5 degrees.

② Run our search process and save search results.

③ Calculate the distance between each search shape and the manually labeled shape.

Since faces in the training images vary widely, as shown in Fig. 2, the Gaussian distribution assumption is unreasonable and the Mahalanobis distance is unreliable. Therefore, ASM collapses in our evaluation and a statistical comparison between ASM and our method is not meaningful. The results of our method are shown in Fig. 3. Compared with the 60 pixels that is the average distance between two eyes of our test faces, the average displacement of all landmarks between every search shape and its manually labeled shape is 2.6 pixels. Average process time of one image is about 2.1 seconds without any optimization to speedup, using Pentium 2.4GHz CPU. Compared to average process time 1.8 seconds of ASM, our method is slightly slow but has a greatly improved robustness.

We also selected images from CMU PIE database [7] to test our method on untrained illuminations. Some examples are shown in Fig. 4. And we testified the robustness of our method on indoor and outdoor photograph taken in backlight or with occlusions. Some results are show in Fig. 5. One example, which has the largest amount of error in our experiments, is shown in Fig. 6. The experimental results show our method is robust to untrained illumination and occlusions on faces, even if we did not use such faces in training.

6 Conclusion

We introduced AdaBoosted histogram classifiers to ASM. Histogram classifiers do not need a Gaussian distribution assumption of local appearances used by previous researches and are very powerful to model big variations of faces. Experimental results demonstrate that our method is robust to changes in illuminations, different expressions, and obstacles. Additionally, unlike previous researches which determine landmark displacements by selecting the most likely position based on Mahalanobis distance, etc., we set displacement to 0 when the highest score in a test profile is lower than 0. This is effective to improve robustness to occlusions on faces.

7 References

[1] T.F. Cootes, C.J. Taylor, "Statistical Models for Appearance for Computer Vision.", http://www.isbe.man.ac.uk/~bim/refs.html, 2004.

[2] Feng Jiao, Stan Z. Li, Heung-Yeung Shum, and D. Schuurmans. "Face Alignment Using Statistical Models and Wavelet Features". Vol.1, Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. pp.321-327. 2003.

[3] R. E. Schapire, Y. Freund, P. Bartlett, andW. S. Lee. Boosting the margin: a new explanation for the effectiveness of voting methods. Ann. Stat., 26(5):1651–1686, 1998.

[4] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," IEEE Conference On Computer Vision and Pattern Recognition, 2000, pp. 746–751.

[5] Paul Viola and Michael Jones, "Robust real-time object detection," Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and sampling 2001.

[6] S. Ioffe, "Red eye detection with machine learning", IEEE Conference on Image Processing, 2003.

[7] T.Sim, S.Baker and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition, 2002.

Fig. 6 Experimental results on photographs. Left: initial states. Right: search results.

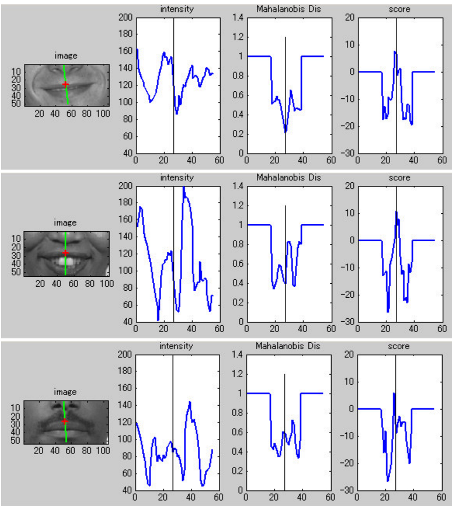

Fig. 2 Column1: cropped image. Green lines show pro-files and red crosses show selected landmarks labeled manually. Column2: intensity distributions of profiles. Column3: Mahalanobis distance distribution of profiles. Column4: AdaBoost score distributions of profiles. Black lines in column2~4 show the positions of landmarks.

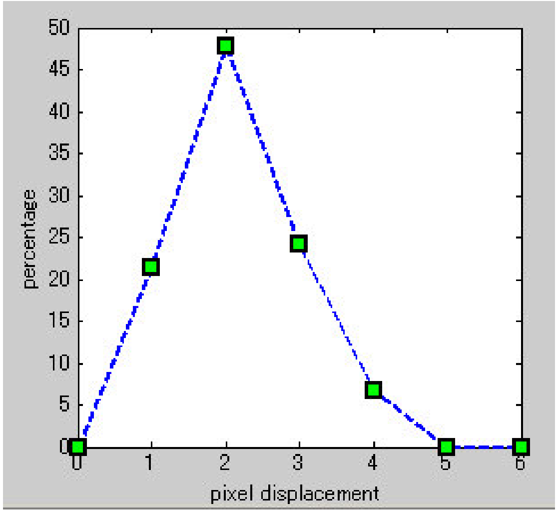

Fig. 3 Point displacement test results. X-axis is the average displacement in pixels. Y-axis is the percentage of points whose displace-ment to the target is X.

Fig. 4 Experimental results on CMU PIE images. Above: initial state. Below: search result.

Fig. 5 Experimental results on photographs. Above: initial state. Below: search result.