

# ROBUST FACE ALIGNMENT BASED ON LOCAL TEXTURE CLASSIFIERS

Li ZHANG<sup>1</sup>, Haizhou AI<sup>1</sup>, Shengjun XIN<sup>1</sup>, Chang HUANG<sup>1</sup>, Shuichiro TSUKIJI<sup>2</sup>, Shihong LAO<sup>2</sup>

<sup>1</sup> Computer Science and Technology Department, Tsinghua University, Beijing 100084, China

<sup>2</sup> Sensing Technology Laboratory, Omron Corporation, Kyoto 619-0283, Japan

E-mail: ahz@mail.tsinghua.edu.cn

## ABSTRACT

*We propose a robust face alignment algorithm with a novel discriminative local texture model. Different from the conventional descriptive PCA local texture model in ASM, classifiers using LUT-type Haar-like features are trained from a large data set as local texture model. The strong discriminative power of the classifier greatly improves the accuracy and robustness of local searching on faces with expression variation and ambiguous contours. A Bayesian framework is configured for shape parameter optimization and the algorithm is implemented in a hierarchical structure for efficiency. Extensive experiments are reported to show its accuracy and robustness.*

## 1. INTRODUCTION

Face alignment, whose objective is to localize the feature points on face images such as the contour points of eye, nose, mouth and outline, is essential to many face processing applications including face recognition, modeling and synthesis. Extensive research has been conducted on face alignment in recent decade. Active Shape Model (ASM) and Active Appearance Model (AAM), proposed by Cootes et al [1], are two popular face alignment methods.

Both of ASM and AAM use a point distribution model to parameterize a face shape with PCA method, but their feature model and optimization strategy are different. ASM method introduces a 2-stage iterative algorithm: 1) given the initial labels, searching for a new position for every label point in its local neighbors, which best fits the corresponding local 1-D profile texture model; 2) interpreting the shape parameters which best fit these new label positions. AAM differentiates itself from ASM with its global appearance model, which is used to directly conduct the optimization of shape parameters. Due to the different optimization criteria, ASM performs more accurately on shape localization, and is relatively more robust to illumination and bad initialization. In this paper, we focus our work on ASM method.

Since ASM only uses a 1-D profile texture feature, which is not sufficient to distinguish feature points from their neighbors, the ASM algorithm often suffers from local minima problem in the local searching stage. Many more representative texture features and more discriminative pattern recognition methods are introduced to reinforce the ASM local searching, such as Gabor wavelet [2], Haar wavelet [3], Ranking-Boost [4] and Fisher-Boost [5], etc. However, face alignment with an accurate local texture model, which can be generalized to large data sets, is still an unachieved goal for practical use.

To overcome the problem, in this paper, we propose a robust ASM method with a novel boosted local texture model learned from a very large set by Real AdaBoost [6] based on LUT-type Haar-like features [8], which has been proved to be very effective in face detection area.

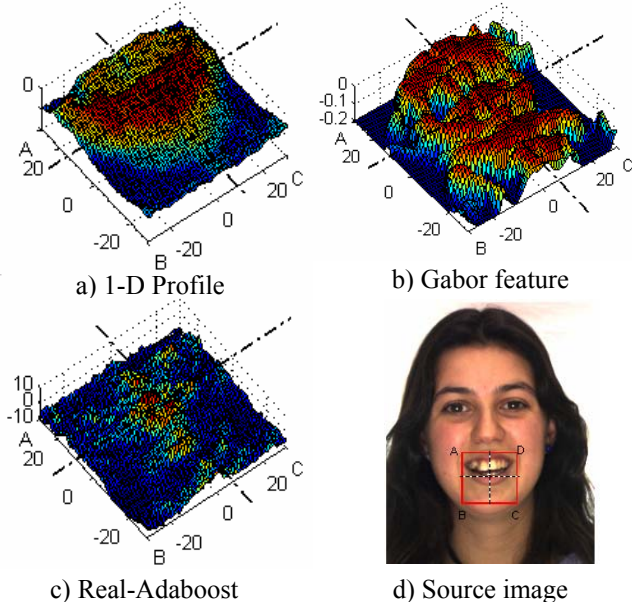
The rest of the paper is organized as follows: In Section 2, local texture model with very strong discrimination power is introduced. In Section 3, a Bayesian framework is configured for shape parameter optimization and is implemented in a hierarchical structure for efficiency. Experiments are reported in Section 4. Finally, conclusions are given in Section 5.

## 2. LOCAL TEXTURE MODEL

The classical ASM method characterizes its 1-D profile texture model normal to the contour with Principle Component Analysis (PCA). However, since PCA didn't consider discriminative criterions between positive samples (feature points) and negative samples (none feature points, its neighbors), the result of local searching stage often falls into local minima.

In order to distinguish feature points from non-feature points, which is critical to diminish the effects of local minima problem, we propose to learn a local texture classifier by boosting a 2-D Haar-like feature over a large training set. This kind of boosted classifier has been proven to be robust and efficient in face detection research.

The LUT-type weak classifiers used here are constructed based on 2-D Haar-like features, which correspond to simple rectangle patches and can be calculated efficiently using the integral image [7]. They



**Figure 1.** (a),(b),(c) shows the confidence by different methods around the feature point of lower-lip (cross point in (d)). (c) successfully distinguishes the feature point(center) from its neighbor while (a) and (b) not discriminative enough.

are extension to original threshold-type classifiers in a finer scale that can greatly improve learning performance [8].

For each feature point, we define the region of interest as a  $[-16,16] \times [-16,16]$  window centered at the feature point. Positive samples are collected from image points within a  $[-2,2] \times [-2,2]$  sub-window at the center, while negatives are sampled randomly out of the sub-window within the region of interest. Then through Real AdaBoost learning [6], sample weights are adapted to select and combine weak classifiers into a strong one which outputs a confidence value with very high discrimination power on the corresponding feature point:

$$conf(\mathbf{x}) = \sum_{t=1}^T h_t(\mathbf{x}) - b$$

In order to illustrate the discrimination power of different methods, the confidence outputs around the feature point of lower-lip as its neighborhood are compared between the classical ASM's 1-D profile feature [1], a Gabor wavelet based feature [2] and our boosted classifier in Figure 1. The origin of (a) (b) (c) corresponds to the cross point of two axis in (d), which is the ground truth position of the feature point. It shows that the boosted feature has a more discriminative confidence than the other two between ground truth position and its neighbors.

Based on the boosted classifier, a local searching procedure will find a maximal confidence around the current label point to be the new feature position. To constrain all the new positions under a global shape model, a shape parameter optimization stage is followed in the next section.

### 3. SHAPE PARAMETER OPTIMIZATION

Suppose label  $S$  is a sequence of the label points in the image  $I$ , which is modeled by parameter  $p$  and  $q$ , as

$$S = T_q(\bar{S} + U \cdot p) \quad (3.1)$$

where  $p$  represents parameter of the point distribution model constructed by PCA,  $\bar{S}$  is the average shape and  $U$  is the matrix of first  $k$  leading eigenvectors;  $Tq(S)$  is the geometrical transformation based on 4 parameters: scale, rotation, and translation.

The objective of shape parameter optimization can be represented in a Bayesian framework [4], as a maximum posterior probability problem:

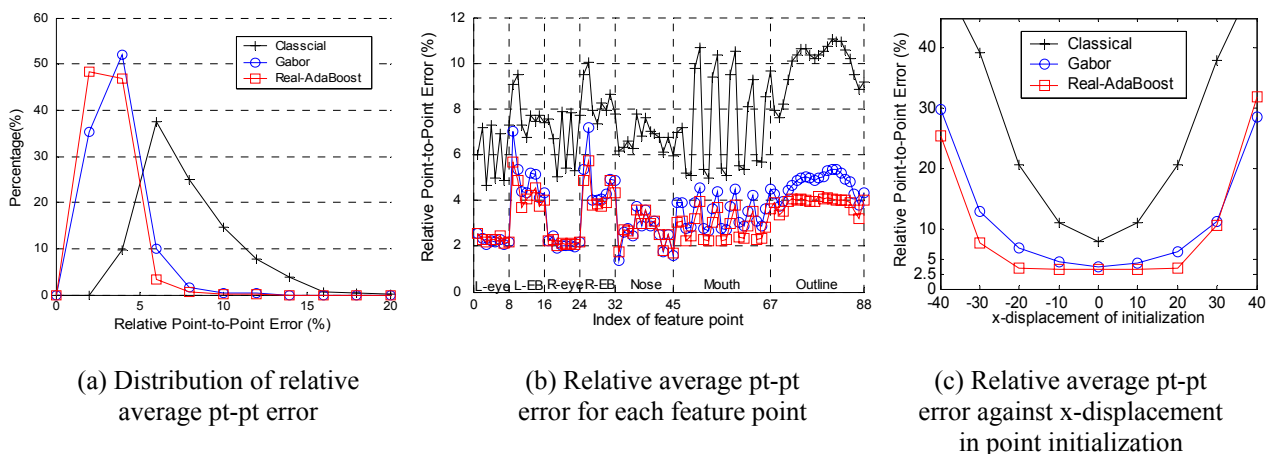
$$\begin{aligned} [p, q] &= \arg \max_{S=T_q(\bar{S}+U \cdot p)} P(S | I) \\ &= \arg \max_{p, q} P(I | S)P(S) \end{aligned} \quad (3.2)$$

where the prior probabilistic distribution in shape parameter subspace  $P(S)$  can be modeled as a Gaussian distribution from PCA construction. And the likelihood of image texture  $P(I | S)$  can further decomposed into independent local likelihood functions of every feature point, which can be assumed to be Gaussian functions centered at the new positions acquired in local searching stage. Thus, the optimization is represented as:

$$\begin{aligned} &\arg \max_{p, q} \prod_i P(I_i | x_i, y_i) \cdot \exp\left(-\alpha \sum_k \frac{p_k^2}{\lambda_k^2}\right) \\ &= \arg \min_{p, q} \sum_i \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} x_i^* \\ y_i^* \end{bmatrix} \right\|_{\sigma}^2 + \alpha \sum_k \frac{p_k^2}{\lambda_k^2} \end{aligned} \quad (3.3)$$

where  $[x_i^*, y_i^*]^T$  is the corresponding local search result.

The optimization (3.3) can be solved with Gauss-Newton method efficiently. However, the Gaussian assumption of local likelihood function may fail in the case of large expression variation and ambiguous contour whose local texture likelihood has multi-peaks (Figure 1 (c)). To extend the range of convergence and improve the computation efficiency, a hierarchical structure similar to that in [9] is adopted. A 3-level pyramid is built by down sampling the image. Starting from the lowest resolution, the face shape represented by only 1/4 feature points is optimized. Sequentially, the result is refined in the higher



**Figure 2. Comparison of classical ASM, Gabor ASM and Real AdaBoost ASM**

and highest resolution, with 1/2 and all feature points respectively.

### 5. EXPERIMENTS

Experiments have been conducted on a very large data set consisting of 2,000 front face images including male and female aging from child to old people, many of which are with exaggerated expressions such as open mouths, closed eyes, or have ambiguous contours especially for old people. The average face size is about 180x180 pixels. We randomly chose 1,600 images for training, and the rest 400 images for testing. The face shape model is made up of 87 feature points, and for each feature point a Real-Adaboost classifier combining 50 weak classifiers is trained. The algorithm was also tested for real-time application on a desktop video from a normal web camera.

For comparison, classical ASM and ASM with Gabor wavelet feature method [2] were implemented and trained on the same training set.

#### 5.1. Accuracy

The accuracy is measured with relative pt-pt error, which is the point-to-point distance between the alignment result and the ground truth divided by the distance between two eyes. The feature points were initialized by a linear regression from 4 eye corner points and 2 mouth corner points of the ground truth. After the alignment procedure, the errors were measured.

The distributions of the overall average error are compared in Figure 2 (a). It shows that our method outperforms the other two. The average errors of the 87 feature points are compared separately in Figure 2 (b). The x-coordinates, which represent the index of feature points, are grouped by organ. It shows that the

improvement of our methods is mainly on feature points of mouth and contour.

#### 5.2. Robustness

To measure the robustness of our method, we initialized the feature points with a -40 to 40 displacement in x-coordinate from the ground truth when average distance between two eyes is about 87 pixels, and the variation of overall average error was calculated and shown in Figure 2 (c).

Additional comparison of alignment results on images with large expression variations from Purdue AR [10] out of the training/testing set are shown in Figure 3.

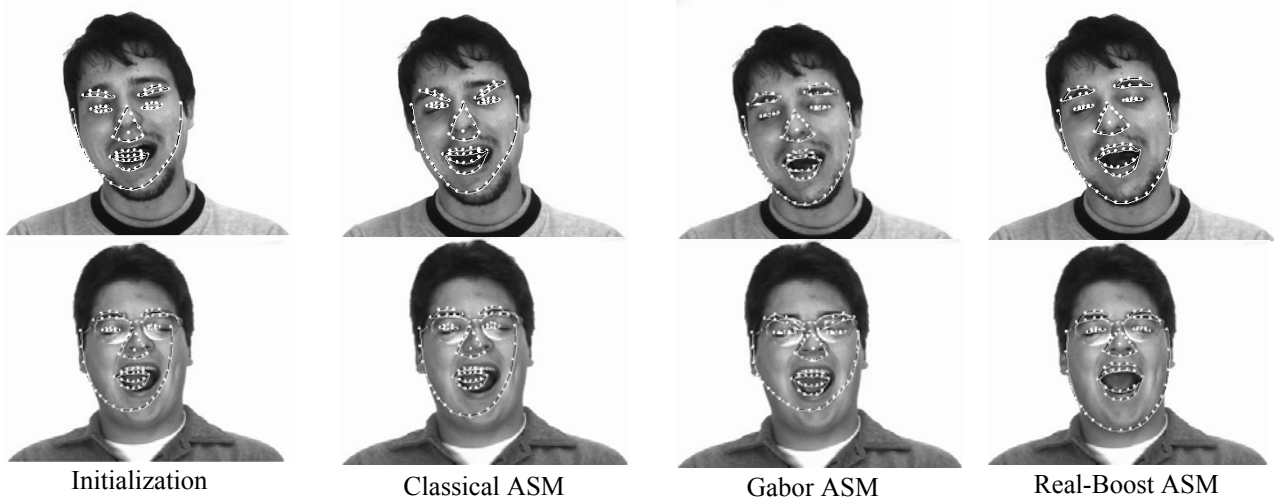
#### 5.3. Efficiency

The average execution time is listed in Table 1. All the tests are carried out on an Intel PIV-3G PC. The classical ASM is the fastest since its computation of local texture model is very simple. Our method is a little slower but is still comparable with the classical ASM. The Gabor-ASM takes much more time.

**Table 1.** The average execution time per iteration (a two-stage process)

Algorithm	Classical	Gabor	Real-AdaBoost
Time per iteration	2ms	567ms	46ms

On smaller faces (100x100 pixels), our method can achieve a real-time performance for desktop face video using a web camera in case that 6 feature points (eye/mouth corners) are manually initialized at the first frame and then on each sequential frame the labels are initialized with the alignment result of the previous frame. The program can run at a speed of about 10 frames per



**Figure 3.** Comparison on images with large expression variation

second, and it is relevantly robust to the variation in pose, see some results in Figure 4.

## 6. CONCLUSION

In this paper, we proposed a robust face alignment algorithm with a novel local texture model. Instead of modeling a local feature descriptively, a classifier is learned from its texture patterns against its neighbor ones as a local texture model. The classifier is of great benefit to the local searching of feature points because of its strong discriminative power. The generality and robustness of the boosting method guarantee the performance especially considering illumination, expression and pose changes. Therefore, compared with existing ones achieving their models in relative small training sets, our method is more promising in practical applications.

## 7. ACKNOWLEDGEMENTS

This work is supported mainly by a grant from OMRON Corporation. It is also supported in part by National Science Foundation of China under grant No.60332010.

## 8. REFERENCES

- [1] T. F. Cootes, *Statistical models of appearance for computer vision*, online technical report available from <http://www.isbe.man.ac.uk/~bim/refs.html>, Sept. 2001.
- [2] Feng Jiao, Stan Li, Heung-Yeung Shum, Dale Schuurmans, Face alignment using statistical models and wavelet features, *Proceedings of IEEE Conference on CVPR*, pp. 321-327, 2003.
- [3] Fei Zuo, Peter H.N. de With, Fast facial feature extraction using a deformable shape model with Haar-wavelet based local texture attributes, *Proceedings of IEEE Conference on ICIP*, pp. 1425-1428, 2004.

- [4] Shuicheng Yan, Mingjing Li, Hongjiang Zhang, Qiansheng Cheng, Ranking prior likelihood distributions for Bayesian



**Figure 4.** Desktop face video tracking

- shape localization framework, *Proceedings of IEEE Conference on ICCV*, pp. 51-58, 2003.

- [5] Jilin Tu, Zhenqiu Zhang, Zhihong Zeng, Tomas Huang, Face localization via hierarchical CONDENSATION with Fisher Boosting feature selection, *Proceedings of IEEE Conference on CVPR*, pp. 719-724, 2004.
- [6] R. E. Schapire and Y. Singer, Improved Boosting Algorithms Using Confidence-rated Predictions, *Machine Learning*, 37, 1999, pp. 297-336.
- [7] P. Viola and M. Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features*, Proceedings of IEEE Conference on CVPR, pp.511-518, 2001.
- [8] Chang Huang, Bo Wu, Haizhou Ai, and Shihong Lao, Omini-directional face detection based on Real-Adaboost, *Proceedings of IEEE Conference on ICIP*, pp. 593-596, 2004.
- [9] Ce Liu, Heung-Yeung Shum, and Changshui Zhang, Hierarchical Shape Modeling for Automatic Face Localization, *Proceedings of ECCV*, pp.687-703, 2002.
- [10] [http://rv11.ecn.purdue.edu/~aleix/aleix\\_face\\_db.html](http://rv11.ecn.purdue.edu/~aleix/aleix_face_db.html)