Multi-View Active Shape Model with Robust Parameter Estimation

Li Zhang, Haizhou Ai
Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
ahz@mail.tsinghua.edu.cn

Abstract

Active Shape Model is an efficient way for localizing objects with variable shapes. When ASM is extended to multi-view cases, the parameter estimation approaches in previous works are often sensitive to the initial view, as they do not handle the unreliability of local texture search, which can be caused by bad initialization or cluttered background. To overcome this problem, we propose a novel algorithm for parameter estimation, using robust estimators to remove outliers. By weighting dynamically, our method acts as a model selection method, which reveals the hidden shape and view parameters from noisy observations of local texture models. Experiments and comparisons on multi-view face alignment are carried out to show the efficiency of our approach.

1. Introduction

Active Shape Model (ASM) introduced by Cootes [2] is a popular method for object registration in images. By learning statistical distributions of shapes and textures from training data, the method can be used to localize objects such as human faces[4, 5, 6, 9, 10, 11, 12].

Many further studies of ASM method have been carried out to improve robustness. These studies have been focused on two aspects. The first is to capture the local texture under non-linear noises by introducing stronger texture descriptors and machine learning methods, such as Gabor [4], Haar wavelet [12], k-NN [8], Ada-Boosting [10], etc. The second is to improve the parameter estimation approach to converge from bad initialization and overcome false local maxima of the local texture models. Li et al. [5] defined additional rules to truncate invalid shapes, while Hill et al [3] and Yan et al. [9] weighted every label differently to overcome local maxima. These improvements have been proved relevantly robust for small and linear shape variations, but they are not capable of dealing with large non-linear variations, especially those caused by 3D view changes and self-occlusion.

Besides, there have been some works which extend ASM to localize objects with large view changes, especially for the application of multi-view face alignment. These works often involve multiple shape and texture models built by non-linear methods such as Gaussian mixtures [1] and Kernel PCA [6], while self-occlusion can also be handled by learning visibility from training data [11]. However, since the texture model used in the local search of each label point depends on the view, these methods are often very sensitive to the estimation of the hidden view parameter. When the initial view is not predicted correct, the results of local search become unreliable. If the estimation of the shape parameter does not deal with the potential outliers, these multi-view ASM approaches will fail.

To overcome this problem, we propose a novel parameter estimation approach for multi-view Active Shape Model. In our method, every label point is weighted dynamically. Only the label points that are consistent with the shape model will have large weights, while outliers have little weights and their influence are eliminated. By doing this, the algorithm acts as a model selection method revealing the hidden view parameter from unreliable observations of label points. Since our method does not completely depend on the local search of each label point, it is more robust against the initial view and cluttered background.

The rest of the paper is organized as follow: The framework of multi-view Active Shape Model is discussed in Section 2. Our novel robust parameter estimation algorithm is introduced in Section 3. Experiments and comparisons are provided to support our idea in Section 4, while the conclusion comes in Section 5.

2. Multi-View ASM Framework

2.1. Classical ASM

In classical ASM, a valid shape denoted by a vector of label points $S = [x_1, y_1, x_2, y_2, \ldots, x_n, y_n]^T$, can be modeled by shape parameters p and q, as

$$S = T_q(\overline{S} + U \cdot p)$$

where p represents PCA coefficients of the point distribution model with average $\overline{S}$ and eigenvectors U; T_q is the geometrical transform using parameter q which includes scaling, rotation, and translation [3].

For each label point, a local texture model in the form of a likelihood function $L_i(I|x_i, y_i)$ is trained to determine if the local texture at (x_i, y_i) of image I matches the texture of i^{th} label point. In order to localize the object in images, ASM procedure goes iteratively [2]:

1. Local search: each label point is updated to (x_i^*, y_i^*) in its local neighborhood by maximizing $L_i(I|x_i, y_i)$.
2. Parameter Estimation: constrain the label points (x_i^*, y_i^*) to a valid shape by finding p, q which minimize

$$\sum_i \left\| [x_i^*, y_i^*]^T - T_q([\overline{x}_i, \overline{y}_i]^T + U_i \cdot p) \right\|^2 \quad (1)$$

where U_i are the two consecutive rows of U corresponding to the i^{th} label.

2.2. Multi-view ASM Framework

There have been some previous works extending ASM to multi-view cases, such as [6, 11]. In this paper, we mainly focus on the parameter estimation approach. Therefore, we use a simplified framework as follow:

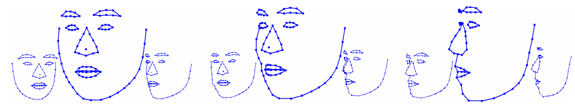
1. Shape models $\overline{S}^v, U^v$ and local texture models $L_i^v(I|x_i, y_i)$ are trained for each view v separately.
2. Given image I, estimate an initial view $v = v_0$.
3. Local search using local texture models $L_i^v(I|x_i, y_i)$.
4. Parameter estimation: find p', q', v' which minimize

$$\sum_i \left\| [x_i^*, y_i^*]^T - T_{q'}([\overline{x}_i^{v'}, \overline{y}_i^{v'}]^T + U_i^{v'} \cdot p') \right\|^2 \quad (2)$$

5. Let the new shape $S = T_{q'}(\overline{S}^{v'} + U^{v'} \cdot p')$, and $v = v'$.
6. if S is not converged go to Step.3.

An example of shape models for multi-view human face alignment is shown in Fig.1.

Compared with previous works, our framework is much simplified. However, like Step.3 of our framework, all previous methods need to specify the view parameter explicitly for local texture search. When the estimation of view is incorrect, the observations from local search are unreliable. To handle this unreliability, we employ a novel parameter estimation approach, which is discussed in Section 3.



(a) View-1 (b) View-2 (c) View-3

Figure 1. Average shapes and variations in their 1^{st} PCA dimensions of 3 face shape models in different views

3. Robust Parameter Estimation

As mentioned in Section 2.2, the parameter estimation approach of Multi-view ASM can be stated as an optimization of equation (2). However, this criterion has two deficiencies.

Firstly, the prior distribution of shape model is not considered. Shapes with very large PCA coefficients should be considered as invalid shape. This constraint can be introduced by penalty terms [9], or truncating coefficients [5].

Secondly, the local search of a label point may fall into a false local maximum. And these outliers will affect the parameter estimation significantly. False local maxima occur when the labels are initialized far from the ground truth or the background is cluttered, which can be common if the initial view is not correct. Weighted methods [3, 9] can partly remove outliers with low likelihoods, but they can not eliminate those with false high likelihoods.

To deal with these two deficiencies, we introduce non-quadratic penalties to remove the effect of outliers. We reform (2) as an optimization minimizing

$$\sum_i c_i^v \rho_i \left(\|e\|^2 \right) + \alpha \sum_i \phi_i(p_i^2) \quad (3)$$

$$c_i^v = \exp(-L_i^v(I|x_i, y_i))$$

$$e = T_q^{-1}([x_i^*, y_i^*]^T) - U_i^v \cdot p - [\overline{x}_i^v, \overline{y}_i^v]^T \quad (4)$$

See [3] for a reference of Equation (4).

$\rho_i(\|x\|^2)$ and $\phi_i(x^2)$ are symmetric, positive-definite penalty functions. When they are quadratic, the solution of (3) reduces to a weighted least-square estimation.

Given parameter v, the p, q minimizing (3) are the solution of following equation,

$$\sum_i w_i A_i^T (A_i [p, q]^T - b_i) + \alpha r [p, q]^T = 0 \quad (5)$$

where

$$\begin{aligned} A_i &= \begin{pmatrix} -U_i^v & \begin{matrix} x_i^* & y_i^* & 1 & 0 \\ y_i^* & -x_i^* & 0 & 1 \end{matrix} \end{pmatrix} \\ b_i &= [\overline{x}_i^v, \overline{y}_i^v]^T \\ w_i &= \exp(-L_i^v(I|x_i, y_i)) \cdot \rho_i'(\|A_i[p, q]^T - b_i\|^2) \\ r &= diag(\phi_1'(p_1^2), \phi_2'(p_2^2), \ldots, \phi_m'(p_m^2), 0, 0, 0, 0) \end{aligned}$$

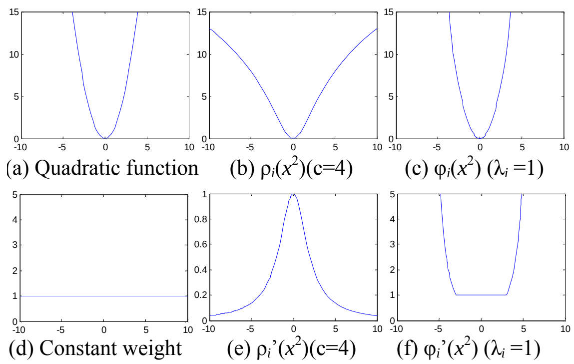


Figure 2. Penalty and weight functions

ρ_i' and ϕ_i' are corresponding first derivatives, and also called weight functions in (5). w_i measures the influence of the observation of each label point, while r represents the constraint on shape parameters. To achieve robustness in the parameter estimation, we should design the penalty functions satisfying following two requirements:

1. ρ_i' should become close to zero when the input error $\|A_i[p,q]^T - b_i\|^2$ is large, so that the effect of these inconsistent observations (outliers) can be eliminated.

2. ϕ_i' should grow significantly when p_i is too large, in order to constrain the shape parameters.

Here, we choose ρ_i and ϕ_i as follow(also in Fig.2):

$$\begin{aligned}
\rho_i(\|x\|^2) &= c^2 \log\left(1 + (\|x\|/c)^2\right) \\
\rho_i'(\|x\|^2) &= 1/\left(1 + (\|x\|/c)^2\right) \\
\phi_i(x^2) &= \begin{cases} x^2/\lambda_i^2, & x \le 3\lambda_i, \\ 9\lambda_i^2 \exp(x^2/9\lambda_i^2 - 1), & x > 3\lambda_i. \end{cases} \\
\phi_i'(x^2) &= \begin{cases} 1/\lambda_i^2, & x \le 3\lambda_i, \\ 1/\lambda_i^2 \cdot \exp(x^2/9\lambda_i^2 - 1), & x > 3\lambda_i. \end{cases}
\end{aligned}$$

where λ_i is the standard derivation of p_i.

From Fig.2, we can see that compared with quadratic function, ρ_i and ϕ_i provide non-constant weight functions. For ρ_i, the output is close to 1 when error is small, but falls to zero when error grows large. Therefore, the solution will be determined only by observations that have small errors, which are consistent with the model. On the other hand, ϕ_i penalizes large values heavily, so that the solution is constrained within the subspace of valid shapes.

By reweighing w_i and r iteratively, we can solve (3) as follow:

- For each view v

  1. let $w_i = \exp(-L_i^v(I|x_i, y_i))$,
     $r = diag(1/\lambda_1^2, \ldots, 1/\lambda_m^2, 0, 0, 0, 0)$.
  2. solve p, q in equation (5) using w_i and r.
  3. reweight w_i and r as
     $$w_i = \exp(-L_i^v(I|x_i, y_i)) \cdot \rho_i'(\|A_i[p,q]^T - b_i\|^2)$$
     $$r = diag(\phi_1'(p_1^2), \phi_2'(p_2^2), \ldots, \phi_m'(p_m^2), 0, 0, 0, 0)$$
  4. If p and q are not converged, go back to Step.2

- Among all the views, select the solution with the minimum of equation (3) as the new shape and view.

Compared with previous works, this algorithm are less sensitive to the initial view, since the outliers are less weighted. Our method not only estimates the shape parameters, but also helps recover the view parameter from cluttered observations. We then apply this algorithm to our multi-view ASM framework to provide robust parameter estimation. The experiments will be discussed in Section 4.

4. Experiments

To illustrate the performance of our approach, we applies our framework to multi-view face alignment, which is to localize a set of facial label points in images.

We use a face database including totally 1800 images taken by a camera array with horizontal pose angles from $-90°$ to $0°$ at $15°$ increments. The size of faces is 250 by 250. For each pose angle, there are 300 images, while we randomly choose 250 for training and others for test.

All the training images are manually labeled with 88 label points. The local texture model we used is the boosted rectangle feature introduced in [10]. 3 pairs of shape and texture models are trained to cover all the pose angles in training data(Fig.1 for an example of the shape models). We implement our multi-view ASM framework with three different parameter estimation approaches for comparison. They are unweighted least-square method [2], weighted least-square method [9] and our robust method.

An illustration of the ASM iterations using different parameter estimation approaches is shown in Fig.3. Even if some labels are not updated correctly by local search due to bad initialization, our approach still recovers the shape and view parameters (Fig.3(d)), while the other two approaches are confused by outliers (Fig.3(e)(f)).

The performance is measured by average point-to-point errors between alignment results and ground truths. Multi-view ASMs using three different parameter estimation approaches are compared in Fig.4. Even though every alignment is initialized with a frontal face shape, we can see that our method still has low errors on full profile images (Fig.4(c)), while the performance of the other two methods becomes poor. It shows that our method is more robust against the initial view. Our approach also outperforms the other two methods on overall performance(Fig.4(d)). Additional test results on CMU-PIE database [7] - which is independent of our training database - are shown in Fig.5.

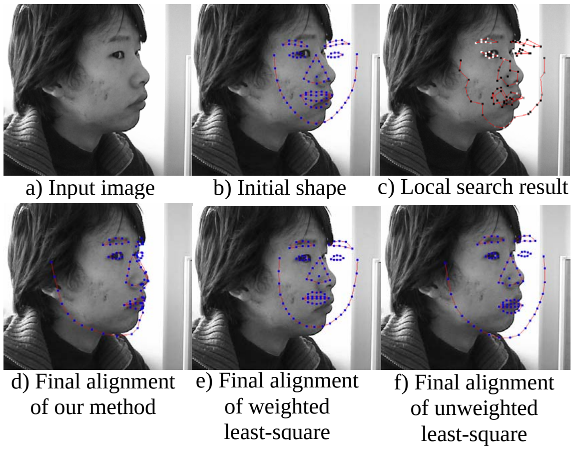


Figure 3. Comparison of parameter estimation approach

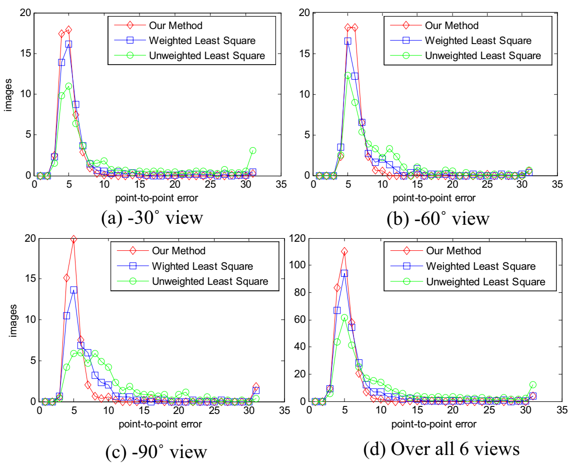


Figure 4. Error distribution of results on different views initialized from frontal faces

Figure 5. Additional results on CMU-PIE

5. Conclusion

In this paper, we have presented a novel method to estimate the shape parameters robustly under a multi-view ASM framework. While outliers are efficiently removed by weighting dynamically, this robust estimator acts as a model selection method which helps reveal the hidden view parameter from unreliable observations. Therefore, compared with previous works, our approach is less sensitive to the initial view and cluttered background. Though our robust parameter estimation has been implemented under a simplified framework, it can also be easily integrated with other multi-view ASM methods to improve their robustness.

6. Acknowledgments

This work is supported by National Science Foundation of China under grant No.60332010.

References

[1] T. F. Cootes and C. Taylor. A mixture model for representing shape variation. In *BMVC*, 1997.

[2] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, http://www.isbe.man.ac.uk/ bim/refs.html, 2001.

[3] A. Hill, T. F. Cootes, and C. J. Taylor. Active shape models and the shape approximation problem. In *BMVC*, 1995.

[4] F. Jiao, S. Li, H. Shum, and D. Schuurmans. Face alignment using statistical models and wavelet features. In *CVPR*, 2003.

[5] Y. Li and W. Ito. Shape parameter optimization for adaboosted active shape model. In *ICCV*, 2005.

[6] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernel PCA. In *BMVC*, 1999.

[7] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression (PIE) database of human faces. the robotics institute, Carnegie Mellon University. Technical report, 2001.

[8] B. van Ginneken, A.F.Frangi, J.J.Stall, and B. ter Haar Romeny. Active shape model segmentation with optimal features. *IEEE-TMI*, 2002.

[9] S. Yan, M. Li, H. Zhang, and Q. Cheng. Ranking prior likelihood distributions for bayesian shape localization framework. In *ICCV*, 2003.

[10] L. Zhang, H. Ai, S. Xin, C. Huang, S. Tsukiji, and S. Lao. Robust face alignment based on local texture classifers. In *ICIP*, 2005.

[11] Y. Zhou, W. Zhang, X. Tang, and H. Shum. A bayesian mixture model for multi-view face alignment. In *CVPR*, 2005.

[12] F. Zuo and P. H. de With. Fast facial feature extraction using a deformable shape model with haar-wavelet based local texture attributes. In *ICIP*, 2004.