Enhance ASMs based on AdaBoost-based Salient Landmarks Localization and Confidence-Constraint Shape Modeling

Zhiheng Niu[1], Shiguang Shan[2], Xilin Chen[1, 2], Bingpeng Ma[2, 3] and Wen Gao[1, 2, 3]

[1] Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin, 150001, China
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China
[3] Graduate School of the Chinese Academy of Sciences, Beijing, 100039, China
{zhniu, sgshan, xlchen, bpma, wgao}@jdl.ac.cn

Abstract. Active Shape Model (ASM) has been recognized as one of the typical methods for image understanding. Simply speaking, it iterates two steps: profile-based landmarks local searching, and statistics-based global shape modeling. We argue that the simple 1D profile matching may not localize landmarks accurately enough, and the unreliable localized landmarks will mislead the following shape matching. Considering these two problems, we propose to enhance ASM from two aspects: (1) in the landmarks local searching step, we introduce more efficient AdaBoost method to localize some salient landmarks instead of the relatively simple profile matching as in the traditional ASMs; (2) in the global shape modeling step, the confidences of the landmark localization are exploited to constrain the shape modeling and reconstruction procedure by not using those unreliably located landmarks to eliminate their negative effects. We experimentally show that the proposed strategies can impressively improve the accuracy of the traditional ASMs.

1 Introduction

In most pattern recognition and computer vision tasks, the localization and alignment of target object from an image is a task of great importance. To deal with the problem, many methods have been proposed in recent years including active contour models (snake) [1], deformable template [2], elastic bunch graph matching [3], Gabor wavelet networks [4], Active Shape Models (ASMs) [5] and Active Appearance Models (AAMs) [6] etc. Among all these methods, ASMs and AAMs, both based on statistical models, has been recognized to be efficient and effective for image interpretation. Simply speaking, ASMs iterate two steps: profile-based landmarks local searching, and statistics-based global shape modeling. In the first step, local texture on the direction perpendicular to the contour, so-called profile, is exploited to model the local texture of each landmark and search for the landmarks locally. The global shape models are then applied in the second step to "correct" the local search result according to the statistical shape model.

The profile-based searching in ASMs provides an efficient way to localize landmarks, which has made ASMs computationally fast. However, due to the variations in pose, lighting, and expressions, local texture of the same landmark may vary dramatically. So, the simple 1D profile model may not localize landmarks accurately enough for further process. Therefore, we argue that some elaborate local texture matching and searching method should improve the final performance of the ASMs, especially for those most significant landmarks. In this paper, AdaBoost based on Haar wavelets which has achieved great success in face detection [7] is introduced to localize some salient landmarks.

Furthermore, in the shape matching step of ASMs, all the landmarks are equivalently treated. However, one can intuitively imagine that some landmarks inherently cannot be located accurately, e.g., some contour points, while others may inherently be easy to localize, e.g., the eyes. It is easy to understand that the imprecisely located landmarks will mislead the following shape matching process. Aiming at this problem, this paper proposes to exploit the confidence of the local matching and reconstruct the shape model according to those reliably localized landmarks with high confidence. Thus, the possibly incorrect landmarks are not involved in the shape matching process.

2 Enhanced ASMs (EASMs)

The system overview of the proposed Enhanced ASMs is illustrated in Figure 1, in which two main contributions can be highlighted: one is the AdaBoost-based salient landmarks location, the other is the shape fitting based on only "reliable" landmarks.

In the following two sections, we will take facial landmarks localization for example to describe these two contributions in detail. Readers can easily generalize it to the analysis of other objects.

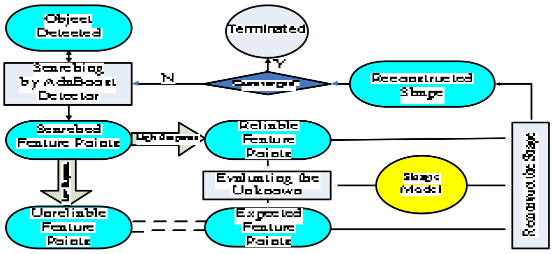

Fig. 1. System overview

2.1 AdaBoost-based Salient Landmarks Localization

The Adaboost we used for salient landmarks localization is quite similar to that for face detection proposed by Viola and Jones [7]. The main difference lies in the adoption of more constraints when preparing the training data, especially the negative (non-landmark) examples, which are all derived from the neighborhood region of the specific facial landmark. This is reasonable considering that we need only to search these landmarks in the face region detected by a previous face detector, such as Viola and Jones' AdaBoost method. For the same reason, in AdaBoost searching, the candidate search region for each landmark is also restricted elaborately. Besides, all the AdaBoost-based landmark detectors can share the same integral image, which can further speed up the procedure.

For face case, we define totally 44 landmarks, 13 contour points and 31 inner points respectively. Since the contour points are intuitively more appropriate to localize by using profile-based local search, only the 31 inner points are localized using the abovementioned AdaBoost-based method.

Note that, in the iteration of the ASM searching, each AdaBoost-based landmark detector naturally provides a detection confidence indicating the "accuracy" of the detection, which will in turn determine whether the corresponding landmark can be used for sequential shape fitting.

Cristinacce and Cootes [8], [9] has also used boosted classifiers similarly, with a Shape Constrained manner, and achieve good experimental result.

2.2 Confidences-Constraint Shape Reconstruction

In traditional ASMs, once the landmarks are locally located in the local search step, a statistical model, pre-trained from a training set based on PCA, is acted on ***all*** the candidate landmarks to find a model best matching with the current candidate shape. This procedure can be formulated as follows:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{Pb} \tag{1}$$

where $\mathbf{x} = \left[x_0, y_0, x_1, y_1, \ldots, x_{n-1}, y_{n-1}\right]^T$ is the shape vector concatenate the coordinates of all the landmarks, $\overline{\mathbf{x}}$ is the mean shape vector, ***P*** is the matrix containing the principle components of the shape vector, and ***b*** is the shape parameters of **x** when projected to the shape subspace.

In this paper, we argue that candidate landmarks with very low detection confidence may decrease the accuracy of the sequential model fitting. So, we propose to utilize only those landmarks with high enough confidence for model fitting.

Firstly, in the current iteration, all the candidate landmarks are categorized into two sets, reliable candidates set (RCS) and unreliable candidates set (UCS), by the following criteria:

$$\begin{aligned} &\mathrm{L_i} \in \mathrm{RCS,if\ C_i} > q_\mathrm{i} \\ &\mathrm{L_i} \in \mathrm{UCS,if\ C_i} \leq q_\mathrm{i} \end{aligned} \tag{2}$$

where $\mathrm{L_i}$ denotes the i^{th} landmark, C_i is the detection confidence of $\mathrm{L_i}$, and q_i is the threshold for $\mathrm{L_i}$. And we assume the number of landmarks in RCS and UCS are r and u respectively. Then, we try to perform PCA by using only the r reliable landmarks, considering the u unreliable landmarks as missing data. Thus, in shape vector **x**, there are $m = 2r$ elements are known and $k = 2u$ elements missing.

We have the transform matrix **P** that will transform the covariance matrix of **x** to diagonal matrix **?** which contains its eigenvalues, and $\mathbf{Q} = \mathbf{P?}^{-1/2}$ that will transform it to identity matrix. Therefore the coordinates in variance-normalized space is denoted as $\mathbf{z} = \mathbf{Q}^T(\mathbf{x}\text{-}\overline{\mathbf{x}})$. There exists a row transform matrix **R**, where $\mathbf{R}^T\mathbf{R} = \mathbf{RR}^T = \mathbf{I}$, that will transform **x** to $\mathbf{x}' = \mathbf{R}^T(\mathbf{x}\text{-}\overline{\mathbf{x}})$ whose first m elements are known and the last k elements are unknown. The objective function we try to minimize is the variance-normalized distance from shape x to the mean shape $\overline{\mathbf{x}}$, it comes from the fact that the shorter the distance is the shapes are more like to be.

$$\mathbf{z}^T\mathbf{z} = \left[\mathbf{Q}^T(\mathbf{x}-\bar{\mathbf{x}})\right]^T \mathbf{Q}^T(\mathbf{x}-\bar{\mathbf{x}}) = \mathbf{x}'^T \mathbf{R}^T\mathbf{Q}\mathbf{Q}^T\mathbf{R}\mathbf{x}'^T \tag{3}$$

It is actually a least square problem. We partition them as $\mathbf{x}' = \mathbf{R}^T(\mathbf{x}-\bar{\mathbf{x}}) = \begin{bmatrix}\mathbf{a}_\mathbf{m}\\ \mathbf{a}_\mathbf{k}\end{bmatrix}$ and $\mathbf{C} = \mathbf{R}^T\mathbf{Q}\mathbf{Q}^T\mathbf{R} = \begin{bmatrix}\mathbf{C}_\mathbf{mm} & \mathbf{C}_\mathbf{mk}\\ \mathbf{C}_\mathbf{km} & \mathbf{C}_\mathbf{kk}\end{bmatrix}$, Where $\mathbf{C}_\mathbf{ij}$ is $i \times j$ matrix and $\mathbf{a}_\mathbf{i}$ is *i*-dimensional column vector. $\mathbf{a}_\mathbf{k}$ is the only independent variable of objective function. Finally we represent the unknown elements (unreliable points) by the know elements (reliable points) and the previous established shape model.

$$\hat{\mathbf{a}}_\mathbf{k} = -\mathbf{C}_\mathbf{kk}^{-1}\mathbf{C}_\mathbf{km}\mathbf{a}_\mathbf{m} \tag{4}$$

The predicted shape vector is:

$$\hat{\mathbf{x}} = \mathbf{R}\hat{\mathbf{x}}' + \bar{\mathbf{x}} = \mathbf{R}\begin{bmatrix}\mathbf{a}_\mathbf{m}\\ \hat{\mathbf{a}}_\mathbf{k}\end{bmatrix} + \bar{\mathbf{x}} \tag{5}$$

Radically our reconstructing of shape by a part of landmarks, reliable points, is a kind of PCA with missing data. We develop this simple but efficient method to solve the problem instead of complicated EM algorithm and Probabilistic Principal Component Analysis [10], [11].

3 Experiments and Analysis

We test the algorithm on two data sets. The first is the face database built by our lab. The images are 240*320, and with 103 manually labeled landmarks. We compare the method introduced in this paper and the conventional ASM on this data set. The second is a publicly available image set known as the BioID database with 20 manually labeled landmarks, used by D. Cristinacce and T. Cootes [9]. We have a comparison with their approach. Our algorithm relies on a global face detector to detect a rectangle of upright front face which is part of the algorithm's input. We use the face detector proposed by Viola and Jones [7].

3.1 Tests on Our Own Database

We test 400 images contained a frontal face on the first data set. We measure the positional error of each feature and express it as a proportion of the inter-ocular distance. The x-axis is the No. of the facial feature points, and the y-axis is the error distance. The mean error of our method is 0.047, while the conventional ASM is 0.084. Our method achieves a good accuracy.

We tests three different methods: the EASM with AdaBoost-based Searching and Confidence-Constraint Shape (ASM_{A+C}), ASM and ASM with AdaBoost-based Searching (ASM_A). Figure 2 shows our method is significantly accuracy and demonstrates that using reliable points to reconstruct shape make the algorithm more robust. This is because the mechanism of reconstructing shape decreases the effect of the improperly searched points to the final shape.

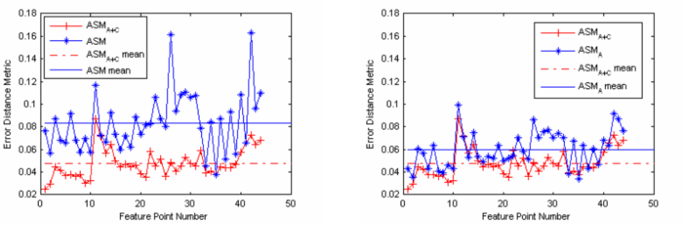

Fig. 2. Average positional error of the 44 feature points comparing with ASM and ASM_A

3.2 Tests on BioID Face Database

The BioID images consist of 1521 images of frontal faces with a set of 20 manually labeled feature points. The images are taken in uncontrolled conditions using a web camera within an office environment. We test our method under the same Testing Criteria described by D. Cristinacce and T. Cootes [9]. The comparison of EASM and SOS is shown in Table 1, where m_{e17} is the mean error distance of the 17 points [9]. Note that the proportion of successful searches of our EASM should be relatively higher if our face detection failures are the same 74 as in SOS, considering that most of the detection failures are low quality images much harder to analyze.

Table 1. Comparison of EASM and SOS [9] using the same 17 feature points

Methods	Global face detection failures	Proportion of successful searches	
		$m_{e17}<0.10$	$m_{e17}<0.15$
EASM	51	84%	97%
SOS[9]	74	85%	96%

Because of the eyes localization is much important in face recognition and some other applications we pay much concentration in eyes searching and localization. Our method achieves high accuracy in eyes localization with 95% of the tested images on BioID data set whose errors are less than 0.1 and in this condition the conventional ASM achieves only 22% (shown in right side of Figure 3).

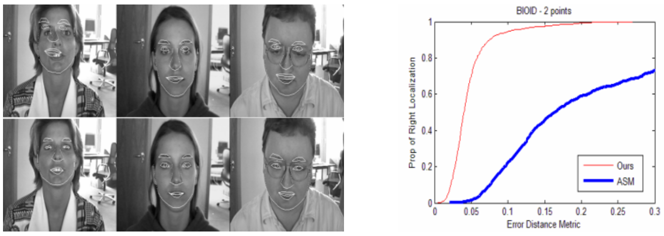

Fig. 3. Some examples of searching results of ASM (up) and EASM (down), and the right side shows average positional error of the eyes, comparing EASM and ASM, on BioID

4 Conclusions and Future Work

In this paper, we present our Enhanced Active Shape Models (EASMs) by using the more accurate AdaBoost-based landmark detector instead of the relatively simple profile matching, and exploiting the confidences of the landmark localization to constrain the shape modeling and reconstruction in order to avoid the negative effects of those unreliably located landmarks. We empirically show that these two improvements can impressively improve the accuracy of the ASMs. Especially, the technique that only the reliable points are used to reconstruct the shape has many advantages: firstly we have decreased the effect of the improperly searched points to the final shape; secondly we can use only part of the landmark points to set the initial shape instead of specific feature points (e.g. eyes), if these feature points are sheltered, it does not necessarily change the final result; thirdly we can easily change the searching strategy and keep the shape model unchanged just by setting the unused feature points invalid.

Currently, the landmarks are directly divided two distinct groups according to their localization confidence, and only the so-called “reliable” landmarks are involved in the shape modeling. However, considering that the confidences are real continuous values, in the future, we will try confidences-weighted shape modeling methods to make better use of the information. How to further expand the proposed methods to AAMs will also be considered in the future.

Acknowledgments

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, "100 Talents Program" of CAS, ShangHai Municipal Sciences and Technology Committee (No.03DZ15013), and ISVISION Technologies Co., Ltd.

References

1. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. International Journal of Computer Vision, pp. 321–331, 1988.
2. A. L. Yuille. Deformable templates for face detection, J. Cogn. neurosci. 3, pp.59–70, 1991.
3. L. Wiskott, J. M. Fellous, N. Kruger, C. Malsburg, Face Recogniton by Elastic Bunch Graph Matching, IEEE TPAMI, Vol.19, No. 7, pp. 775–779, 1997.
4. V. Kruger and G. Sommer. Gabor wavelet networks for object representation. Technical Report 2002, Institute of Computer Science, University of Kiel, 2000.
5. T. F. Cootes, C. J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. CVIU, 61(1), pp. 38–59, January 1995.
6. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In 5th ECCV, volume 2, pp.484–498. Springer, Berlin, 1998.
7. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. CVPR 2001, volume 1, pp.511–518, Kauai, Hawaii, 2001.
8. D. Cristinacce and T. F. Cootes. Facial feature detection using AdaBoost with shape constraints. In 14th BMVC, pp. 231–240, 2003.
9. D. Cristinacce and T. F. Cootes. A Comparison of Shape Constrained Facial Feature Detectors. Proceedings of the Sixth IEEE FGR, Seoul, Korea, 2004.
10. Haifeng Chen. Principal Component Analysis with Missing Data and Outliers. Technical Report 2002, Electrical and Computer Engineering Department Rutgers University, Piscataway, NJ, 08854.
11. M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. Journal of the Royal Statistical Society, Series B, 61, Part 3, pp. 611–622, 1999.