

An Experimental Study on Automatic Face Gender Classification

Zhiguang YANG¹, Ming LI², Haizhou AI¹

¹ Computer Science and Technology Department, Tsinghua University, Beijing 100084, China

² Nanshan Branch, Shenzhen Public Security Bureau, Shenzhen 518052, China

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

Abstract

This paper presents an experimental study on automatic face gender classification by building a system that mainly consists of four parts, face detection, face alignment, texture normalization and gender classification. Comparative study on the effects of different texture normalization methods including two kinds of affine mapping and one Delaunay triangulation based warping as preprocesses for gender classification by SVM, LDA and Real Adaboost respectively is reported through experiments on very large sets of snapshot images.

1. Introduction

Over the past decade, face image processing has achieved very significant advances especially in face detection and face alignment areas that are somewhat matured being able to provide fast and robust algorithms [1][2][3] for practical applications. With the location and shape of face being extracted very accurately, it is natural to consider what further works can be done in the area of face recognition, facial expression recognition, as well as category recognition such as gender, age, ethnicity classifications. In this paper, we focus on the problem of automatic gender classification that separates his faces from her faces. Gender classification could be of important value in human-computer interaction, such as personal interaction.

Earlier work on gender classification mainly originated in psychology and cognition researches [4][5][6]. More recently, people began consider this problem more technically from statistical learning point of view on large data sets, among which representative works are Moghaddam and Yang's RBF-kernel SVMs method [7] that achieved very good results (only 3.4% error) on FERET database and Shakhnarovich et al.'s Adaboost method [8] that achieved even better performance than SVMs. Other work includes Wu et al.'s

LUT-based Adaboost method [10] that implemented a real-time gender classification system with comparative performance. In this paper, we report a comparative study on the effects of different texture normalization methods including two kinds of affine mappings and one Delaunay triangulation based warping as preprocesses for gender classification by SVM, FLD and Real Adaboost respectively through experiments on very large sets of snapshot images.

The rest of this paper is organized as follows: Gender classification system overview is given in Section 2; in Section 3, three texture normalization methods are introduced; in Section 4, the gender classification algorithm is described; the experiment results are shown in Section 5; and the summary is given in Section 6.

2. System overview

The automatic face gender classification system consists of face detection, face alignment, texture normalization and gender classification modules as illustrated in Figure 1. It uses Huang et al. [3]'s face detection module and Zhang et al. [2]'s face alignment module. This paper mainly discusses texture normalization and gender classification modules.

First, face detection module detects the face in the given image. Second, face alignment module aligns 88 facial landmarks, lying on face contour, eyebrows, eyes, nose and mouth, as shown in Figure 2.c. Comparing with face detection, face alignment provides finer position and size of face, and additional shape as well. Then normalization module is used to alleviate the variations due to shape and pose changes, such as in-plane rotation and so on based on the face alignment result. Finally, gender classification module makes decision.

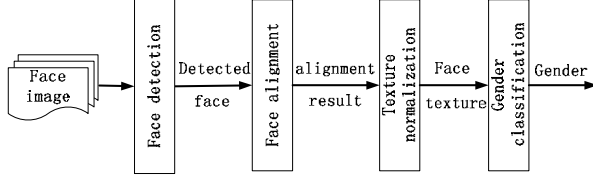


Figure 1. Face Texture normalization

3. Face texture normalization

After face alignment, there are basically three methods to do texture normalization, see Figure 2 for a reference: the first one is to warp face texture based on Delaunay triangulation to a reference neutral mean shape that results in shape-free texture as in AAM [10], the second one is to do an affine mapping fitted on the hypothesis that those 65 facial points on organ contour out of 88 points (excluding face contour points and nose tip) are generally coplanar, the last one is a simplified form of the second by doing an affine mapping fitted between six facial feature points of four eye-corners and two mouth-corners.

Method 1 normalizes texture by triangulation warping, that is to warp each triangle texture of a given face to that of the mean shape. The individual shape is discarded while making the face looks like a mean shape. From another point of view, this normalization method brings more consistence of each facial landmark; it does normalize not only the face size and in-plane rotation, but also the face shape. *Method 2* regularizes the face by affine transformation. Given the coplanar facial points, the size and shape can be estimated well by MSE method. Although shape is not normalized, affine mapping succeeds in maintaining local texture characteristic. *Method 3* is a simplified version of *Method 2*, six points is usually considered to be enough for estimating the size and small pose changes. When face alignment is achievable, it is reasonable to think that *Method 3* may not be so robust as *Method 2*; however, six points extraction doesn't rely on face alignment that can be directly extracted much faster than face alignment approach, which is very common in current face processing research including face recognition etc.

The above three normalization methods may have subtle effects on the performances of succeeding classification methods that will be reported in the following sections.

In practice, finally each face is normalized with an average of 128 and a standard deviation of 64; and the texture size is scaled to 32*32 before fed in gender classifiers.

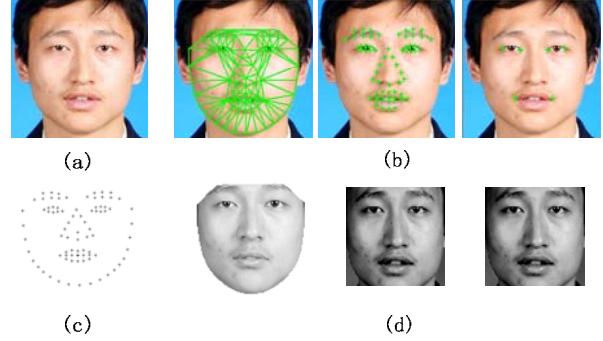


Figure 2. Texture Normalization
(a) input face image; (b) selected landmarks;
(c) mean shape; (d) normalization results

4. Gender classification algorithm

Gender classification is a typical binary classification problem, we implement three different algorithm: SVM, FLD and Real Adaboost.

4.1 SVM

SVM tries to find an optimal hyper plane for linear separable problem. Given M training samples $S = \{(x_i, y_i)\}$, in which $x_i \in R^N$ and $y_i = \pm 1$, the hyper plane could be denoted as $w \cdot x + b = 0$, where w and b are normal vector and bias respectively. And the optimal separating function can be written as:

$$f(x) = \text{sgn}\{(w \cdot x) + b\},$$

For linearly non-separable data, kernel function maps the input sample to a higher dimensional feature space where a linear hyper plane can be found. The kernel based classifier can be written as:

$$f(x) = \text{sign}\left(\sum_{i=1}^m y_i \alpha_i k(x, x_i) + b\right),$$

where $k(\cdot, \cdot)$ is a kernel function. In this paper, RBF kernel function is used.

4.2 FLD

FLD aims at finding the optimal linear projection for classification. Given the sample set in the above section, the discriminate function can be written as

$$f(x) = \text{sign}(w^* \cdot x - b),$$

where w^* is the vector maximizing the ratio between the between-class scatter S_b and within-class scatter S_w and b is the threshold for classification.

4.3 Real Adaboost

Real Adaboost [11] is a statistical learning algorithm by maximizing classification margin iteratively. In each iteration t , one weak classifier $h_t(x)$ is selected from a large hypothesis space; after the final iteration T , all the weak classifiers are combined together to construct a strong classifier. The strong classifier could be written as:

$$H(x) = \text{sign}\left(\sum_{t=1}^T h_t(x) - b\right),$$

where b is an empirical threshold.

Combining with Haar-like feature via integration image, Real Adaboost is effective and fast. And LUT-based [9] weak classifiers can make Haar-like feature more expressive.

5. Experiment

We carry experiments on a database of snapshot images that consists of 11500 Chinese snapshot images, including 7000 male and 4500 female. These faces are always upright and neutral without beard or strange hair style, and some faces keep glasses on. We test our system by five-fold cross validation.

First of all, all the faces are detected and aligned automatically. Then we normalize the face texture by three different methods described in section 3 respectively. Three different classification methods (SVM/FLD/Adaboost) are applied and the average results of five-fold cross validation are shown in Table 1. For the SVM method, PCA features of 95% energy are used. Particularly for *Method 1* of SVM method, two types of feature are used respectively, one is PCA features of shape free texture and the other is concatenated PCA features of shape free texture with shape PCA features (called “appearance”), where shape PCA features are multiplied by a ratio of 10^4 to balance the scale comparable with texture PCA features.

The trained SVM classifiers are with about 1200 support vectors, which is approximately 10% of the training set. The Adaboosted classifiers are with a two-layer structure shown in Figure 3 that are trained in the way that at the first layer it deals with balanced training samples for a coarse classification by an Adaboosted strong classifier and at the second layer each of its two branches deals with unbalanced training samples by an Adaboosted cascade classifier [1] in which the dominant class is treated as negative. In the experiment, the first layer strong classifier is composed of 15 weak classifiers with an error rate of 7.5%; in each branch of the second layer, the cascade classifier consists of two nodes of which the first rejects about

80% negative samples that is composed of 35 weak classifiers, and the second rejects about 90% negative samples further that is composed of 55 weak classifiers.

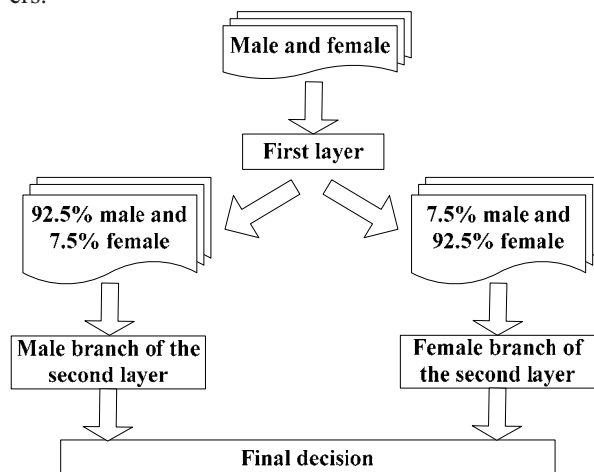


Figure 3. A two-layer structure of real Adaboost classifier

Table 1. Average error rate of five-fold cross validation

	<i>Method1</i>		<i>Method2</i>	<i>Method3</i>
	Shape free texture	Appearance		
SVM	3.25	2.84	2.90	2.95
FLD	3.77	3.01	3.65	3.89
Real Adaboost	3.13	-----	3.54	3.58

The experimental result in Table 1 shows that all three classification methods have achieved less than 4% error rates under all three texture normalization methods, that is to say all three preprocessing methods are effective. Further, it can be seen that no particular texture normalization method is significantly better. Among the three normalization methods, shape free texture achieves the best performance on Adaboost that is reasonable since it aligns the texture best. However shape free texture is not as good as the other two methods on SVM or FLD, but when shape free texture is combined with shape feature to form appearance feature, it do outperform the other two.

Although we focus on Chinese (yellow race) snapshot images, it would be interesting to know its generalization ability to other database. Here in Table 2, we give test results corresponding to shape free textures method on the FERET dataset of which only 3529 frontal FERET images are tested directly using the above mentioned classifiers. In comparison, a five-fold cross validation experiment is also given in Table 3. It is reasonable to achieve better performance since the data are adapted by retraining on those data.

Table 2 Testing error rate on frontal FERET dataset

Method	SVM	FLD	Real Adaboost
Error rate	15%	17.7%	8%

Table 3 Average error rate of five-fold cross validation on frontal FERET dataset.

Method	SVM	FLD	Real Adaboost
Error rate	7.8%	14.3%	6.2%

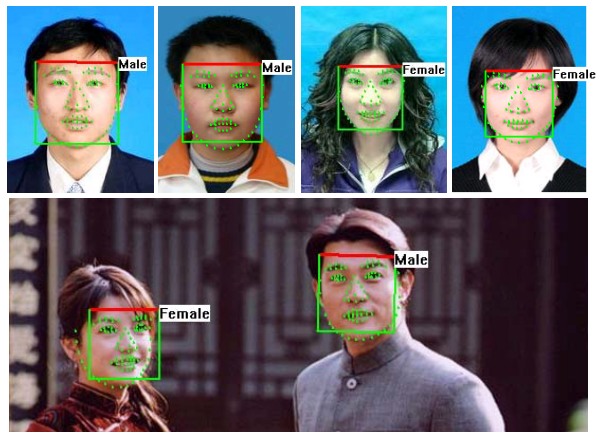
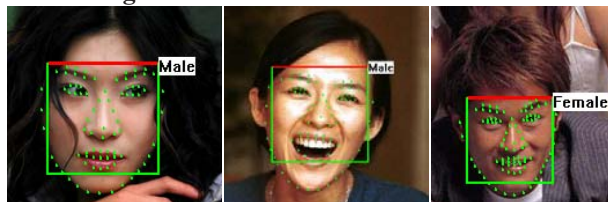
For a discussion, we know that Real Adaboost uses Haar-like local features of face, while SVM and FLD rely on ensemble face template textures. In this sense, we may conclude that when local features are used, shape free texture may be a better normalization method since it is more coherent by excluding shape effects; but when global features are used, fine shape seems to have delicate effects on discriminative power so the other two affine alignment methods work better. As for the fact that the affine mapping based on 65 facial landmarks is a little better than that based on 6 facial landmarks, it may be because of more points contributing to better approximation of affine transformation. But since 6 facial landmarks can be extracted faster compared with face alignment procedure, it is much desirable in many practical applications. As for the speed of computation, the SVM method takes about 5.2ms for a 32x32 candidate image patch, the FLD method takes less than 0.1 ms, and the Adaboost method takes about 0.26 ms. For illustration, some experiment results are given in Figure 4. Some failures are shown in Figure 5 since it is sensitive to expression change and strange hair style.

6. Summary

In this paper, we present an experimental study on automatic face gender classification that is focused on the effects of three different preprocessing methods to the performances of gender classification by SVM, LDA and Real Adaboost. Normalization preprocessing is quite effective for gender classification. In particular, normalization by Delaunay triangulation, which results in shape free texture, is proven to be good at enhancing coherence of local features, and therefore it is suitable for Haar-like feature based Adaboost approach; normalization by affine mapping is better for global features based method such as FLD or SVM method of using ensemble face texture features. As for the simplified six-points based affine mapping, it is more desirable due to its possibility of being extracted fast that is independent of a complex face alignment approach.

7. ACKNOWLEDGMENTS

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

**Figure 4.** Gender classification results**Figure 5.** Failure results of gender classification

8. References

- [1] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", CVPR 2001.
- [2] Li ZHANG, Haizhou AI, et.al, Robust Face Alignment Based on Local Texture Classifiers, ICIP 2005.
- [3] Chang HUANG, Haizhou AI, et.al, Boosting Nested Cascade Detector for Multi-View Face Detection, ICPR 2004.
- [4] A. Golomb, D. T. Lawrence, and T. j. Sejnowski. SEX-NET: A neural network identifies sex from human faces. In *Advancds in Neural Information Processing Systems*, pp. 572-577, 1991
- [5] G. w. Cottrell and J. Metcalfe. EMPATH: Face, emotion, and gender recognition using holons. In *Advances in Neural Information Processing Systems*, pp. 564-571, 1991
- [6] Alice J. O'Toole et al. The Perception of Face Gender: The Role of Stimulus Structure in Recognition and Classification. *Memory and Cognition*, Vol. 26, pp. 146-160, 1997
- [7] B. Moghaddam and M. H. Yang. Gender Classification with Support Vector Machines. *IEEE Trans. On PAMI*, Vol. 24, No. 5, pp. 707-711, May 2002.
- [8] G. Shakhnarovich, P. Viola and B. Moghaddam. A Unified Learning Framework for Real Time Face Detection and Classification. *IEEE conf. on AFG* 2002.
- [9] Bo Wu, Haizhou Ai, Chang Huang, LUT-Based Adaboost for Gender Classification, In *AVBPA' 2003*, Vol.2688, pp.104-110.
- [10] T. F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance model, *ECCV* 1998, vol. 2, pp. 484-498.
- [11] R. E. Schapire and Y. Singer, Improved Boosting Algorithms Using Confidence-rated Predictions. In: *Machine Learning*, 37, 1999, 297-336.