A Hybrid Approach to Gender Classification from Face Images

Ziyi Xu, Li Lu and Pengfei Shi
Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China, 200030
{seiya_xu, lulihappy, pfshi}@sjtu.edu.cn

Abstract

Recently, gender classification from face images has attracted a great deal of attention. It can be useful in many places. In this paper, a novel hybrid face coding method by fusing appearance features and geometry features is presented. We choose Haar wavelets to represent the appearance features and use AdaBoost algorithm to select stronger features. Geometry features are regarded as apriori knowledge to help improve the classification performance. In this work, Active Appearance Model (AAM) locates 83 landmarks, Thus we can get 3403 geometry features, from which 10 most significant features are picked, normalized and fused with the appearance features. Experimental results show the effectiveness and robustness of the proposed approach regarding expression, illumination and pose variation in some degree.

1. Introduction

Gender Classification is a binary classification problem, in which one has to predict an image as that of a man or woman. It is an easy job for human beings, but a challenging one for computers.

With the evolution of Human-Computer Interaction technology (HCI), to meet the people's growing demand for secure, reliable, convenient and individualized services, computer vision approaches such as face identification, gesture recognition, and perhaps most fundamentally gender classification will play an increasingly important role in our daily lives.

In the early stage, gender classification mainly focused on psychology and cognition regions [1, 2]. In recent years, people began considering this problem more technically. Representative works includes Costen et al.'s sparse SVM gender classifier which achieved very good result (94.42% classification rate with Japanese face images) [3] and Shakhnarovich et al.'s automatic Adaboost demographic analysis system that can achieve even better performance than SVMs[4]. Other work includes Wu et al.'s LUT-based Adaboost method with comparative performance [5].

Based on the type of features used, previous studies can be broadly classified into two categories: appearance feature-based (global) and geometric feature-based (local). The former finds the decision boundary directly from training images while the latter is based on geometric features such as eyebrows thickness, nose width, etc. The first strategy usually considers an image as a high-dimensional vector and extracts features from its statistical information, without relying on knowledge about the object of interest. Typically, this approach is holistic and has the advantage of being fast and simple. However, such a representation can become unreliable when local appearance variations occur. In the latter strategy, some apriori knowledge was applied, and facial geometry features such as the eyes, nose, and mouth are extracted first. Therefore, this approach has the advantage of translation and rotation invariability. But it may throw away a lot of information that can be helpful. Ref. [6] presented that recognition difficulty was statistically linked to the type of feature extraction under different subject covariate factors such as age and gender. Its results show that local and global features supplement each other under some conditions. Hence, the hybrid method using both local and global features is presented of itself.

2. Method

Our hybrid gender classification method consists of three main modules: the first one normalized a given face image (some geometry alignment and gray level normalization are processed in this stage), the second one extracts features from the normalized image to

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

form a feature vector, and the third one is the uses this vector as the input, the output of the module reaches the conclusion. The architecture is illustrated in Fig. 1. We will explain carefully in the following passage.

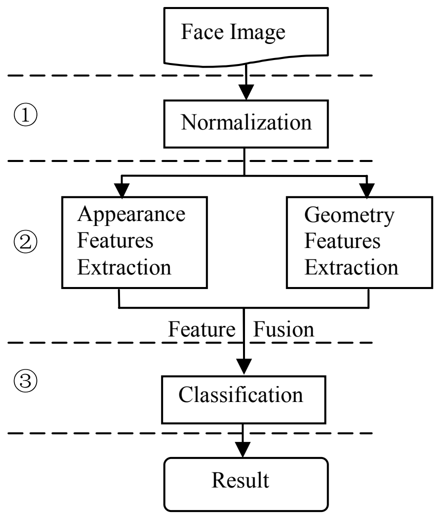


Fig.1 Architecture of gender classifier.

2.1. Feature extraction

Our brain can discriminate thousands of patterns. But we have very few linguistic terms that allow us to verbalize which are the characteristics our brain uses to perform the discrimination. On one hand, restricting ourselves to the features which must have a physical meaning restricts our possibilities of finding much more other useful features. For example we can choose a pattern on the basis of their statistical properties. On the other hand, usually the performance may be improved a lot if we take advantage of the apriori knowledge about the images. In this paper, the appearance features are chosen mathematically by AdaBoost algorithm and the geometry features are extracted using AAM (Active Appearance Model).

2.1.1 Extraction of global features. There are many algorithms proposed for the extraction of global features. In this paper, we choose AdaBoost algorithm to extract Rectangular features from normalized facial images. AdaBoost algorithm, which is the short form for Adaptive Boosting, is a machine learning algorithm. It was first formulated by Freund and Schapire in 1995 [7]. In its original form, the AdaBoost algorithm is used to boost the classification performance of a simple classifier. It does this by combining a collection of weak classifiers to form a stronger one. Usually, the weak classifiers have the advantage of easy construction and low cost of computation.

In our experiment, we choose Haar-like features introduced by Viola et al. (see Fig. 2) [8, 9] and improved by Lienhart. et al. (See Fig. 3) [10] to be the weak classifiers. Each rectangle feature computes the sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the black rectangles. Within any image sub-window the total number of Rectangular features is very large. Drawing an analogy between weak classifiers and features, the AdaBoost algorithm is an effective procedure for searching out strong features. It excludes a lot of "useless" features, and focus on a small set of "useful" features, which contains much more gender information than the others.

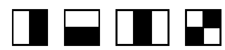


Fig. 2 Rectangle Features used by Paul Viola

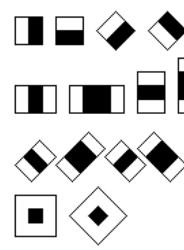


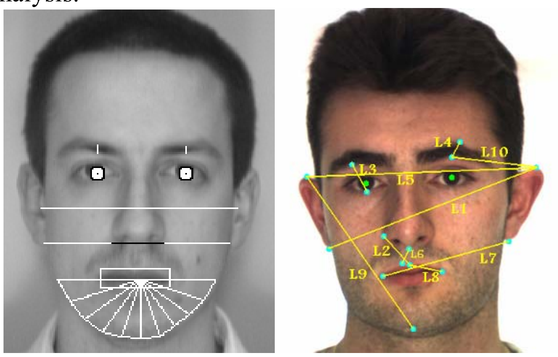
Fig. 3 Extended Features used by Rainer Lienhart

The result class assigned to the input corresponds to the sign of $H(x)$.

$$H(x) = \left(\sum_{j=1}^{T} \alpha_j h_j(x) - b \right)$$

Where $h_j(x)$ is a weak classifier, T is the number of weak classifiers, and b is an empirical threshold.

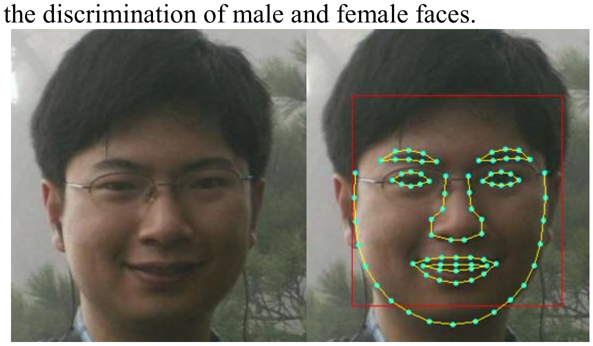
2.1.2 Extraction of local features. There also have been some algorithms for extraction of local features. In Ref. [11], R. Brunelli and T. Poggio pointed out that among the sixteen geometry features they selected for the classification system, the following three features: distance of eyebrow from eyes, eyebrows thickness and nose width, contributed most noticeably, as illustrated in Fig. 4(a). Their correct classification rate reached 79%. Ashok Samal et al. picked mush more features than R. Brunelli and T. Poggio. They measured 406 geometry features. They used statistical analysis to demonstrate that sexual dimorphism does exist in the human face [12] and noted that around 85% of features show significant difference in male and female features. Specially, they pointed out that it is possible to obtain a high gender classification rate of 96% correct classification using 18-20 geometry features. This conclusion lends strong support to the idea that geometry features can be regarded as apriori knowledge for gender classification. Fig. 4(b) shows the ten most significant features that Ashok Samal et al. found out identified by step-wise discriminant analysis.



(a) **(b)**

Fig. 4 Examples of geometry features.

In our experiment, we use AAM to get 83 landmarks from a face image, as shown in Fig. 5(a, b). We can get C_{83}^{2} = 3403 distances. We can get the first ten steps of step-wise discriminant analysis [12]. It lists ten features which contribute most significant towards the discrimination of male and female faces.



(a) **(b)**

Fig. 5 Landmarks located by AAM.

2.2. Fusion of Global and Local Features

There are several normalization methods, e.g. min-max (MM), z-score (ZS), tanh (TH) and adaptive (AD). The study of Ref. [13] shows that the MM and AD methods lead to better performance than other methods. What is more, the MM method is quite simple and easy to calculate. Therefore the MM method was applied in this paper. This method maps the raw scores to the [-1, 1] range[1]. The quantities max(S) and min(S) specify the end points of the range:

$$n = \frac{2 \times S - \max(S) - \min(S)}{\max(S) - \min(S)}$$

[1] http://www.faqs.org/faqs/ai-faq/neural-nets/part2/

3. Image database

The experiment is carried out on a set of 14756 pictures *2* (8544 males and 6212 females). These images are frontal, with some variations in pose, expression, accessory and illumination condition. They were obtained mainly from the following databases: FERET [14], AR [15], and a web picture database serves as a complement. A subset with 12,000 images, of which 7,000 are male and 5,000 are female faces, is selected arbitrarily to train an appearance features extraction system. Another subset with 200 images selected randomly is used to train the geometry classifiers. Each sample is normalized both in geometry and in illumination. The alignment in geometry is based on two eye centers detected automatically by AAM. After determining the face area with the alignment algorithm, the areas were ascaled to the size of 40x44 pixels. Histogram equalization was applied to the extracted face images to normalize for different lighting conditions. Typical samples after normalization are shown in Fig. 6.

Fig. 6 Samples of face images used in training Haar-like features extraction system.

We choose 1000 pictures (500 males and 500 females) at random from the remaining 2756 pictures for experiment via a 5-fold cross-validation.

4. Experiment and results

The dataset was randomly divided into 5 subsets, with each subset having 800 (400 males and 400 females) for training and 200 (100 males and 100 females) for testing. The test sets were not overlapped with their respective training sets and other test sets.

We used a SVM[3] with the radial basis function (RBF) kernel as the classifier. We also tried other kernels like linear kernel and polynomial function, but found RBF kernel produced better results. Since we consider two kinds of feature extraction methods and fuse them together, the classification performance of different combination patterns is analyzed here. As can be seen from Fig. 7. with more and more geometry features taken into consideration, the classification performance of 83% rose steadily and significantly.

[2] The number is doubled by adding the mirror.

[3] http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html

The peak point of the recognition rate of hybrid method was achieved when the eighth geometry feature was added.

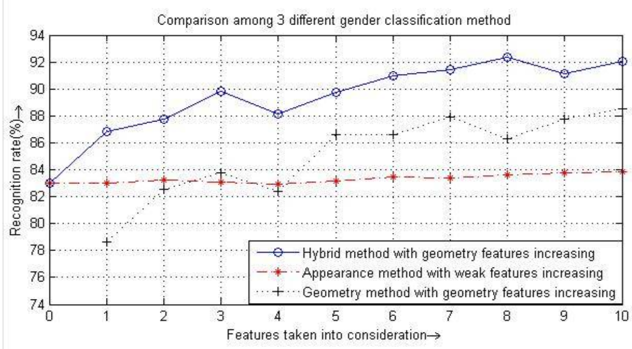


Fig. 7 Performance analysis

Table 1 shows that the performance of the proposed hybrid method via a 5-fold cross-validation. As can be seen, it outperforms the unmixed appearance- or geometry-feature based methods quite a lot.

Recognition rate	Recognition method		
	Appearance only (with 143 weak classifiers)	Geometry only(10 features)	Hybrid method
Total rate	82.97%	88.55%	92.38%
Female	82.19%	89.12%	92.53%
Male	83.74%	87.98%	92.21%

Table 1 Recognition rate of different methods

5. Conclusion and Future Work

In this paper, a new hybrid method of fusing global features and local features was presented. We use AdaBoost algorithm to extract global information and combined with local features which are extracted by AAM. Experimental results demonstrate that the hybrid method obtains a higher accuracy. Moreover, the novel classification method shows effectiveness and robustness regarding expression, illumination and pose variation in some degree.

There are still some aspects that deserve further study. First of all, the AAM algorithm need to locate 83 points for each face, but only a small part of them is useful to form a geometry feature vector. The time consumed for location and feature extraction can be decreased if a further optimization is made.

References

[1] A. Golomb, D. T. Lawrence, and T. j. Sejnowski. "SEXNET: A neural network identifies sex from human faces", *Neural Information Processing Systems*, 1991, pp: 572-577

[2] G. w. Cottrell and J. Metcalfe. EMPATH: "Face, emotion, and gender recognition using holons". *Neural Information Processing Systems*, 1991, pp: 564-571

[3] Costen, N., Brown, M., Akamatsu, S., "Sparse models for gender classification". *Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'04)*, 2004, pp: 201–206..

[4] Shakhnarovich, G., Viola, P.A., Moghaddam, B., "A unified learning framework for real time face detection and classification", *Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'02)*. IEEE, 2002, pp: 14–21.

[5] B. Wu, H. Ai, C. Huang, "Real Time Gender Classification", *Multi-spectral Image Processing and Pattern Recognition*, 2003.

[6] Givens G, Beveridge J R, Draper B A, et al. "How features of the human face affect recognition: a statistical comparison of three face recognition algorithms", *IEEE computer society conference on computer vision and pattern recognition*, 2004, 2 pp: 381-388.

[7] Freund Y., Schapire R. E., "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *Journal of Computer and System Sciences*, 1997, 55(1) pp: 119-139.

[8] Viola P., Jones M. J. "Robust Real-Time Face Detection", *International Journal of Computer Vision* , 2004, 57(2) pp: 137-154

[9] Viola P., Jones M. J., "Rapid Object Detection using a Boosted Cascade of Simple Features", Computer Vision and Pattern Recognition, 2001, Volume 1, pp: 8-14

[10] R. Lienhart, A. Kuranov, V. Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection", *MRL(Microprocessor Research Lab, Intel Labs) Technical Report*, 2002.

[11] R. Brunelli and T. Poggio. "Hyperbf networks for gender classification", *DARPA Image Understanding Workshop*, 1992, pp: 311–314.

[12] Ashok Samal *, Vanitha Subramani, David Marx, "Analysis of sexual dimorphism in human face", J. Vis. Commun. Image R. 18 , 2007, pp: 453–463

[13] Snelick R, Uludag U, Mink A, et al. "Large scale evaluation of multimodal biometric authentication using state-of the-art systems" , *IEEE Transactions on Pattern Analysis and Machine intelligence2005*, pp: 450-455

[14] Phillips, P.J., H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithm", *Image and Vision Computing*, 1998, vol. 16, no. 5, pp. 295-306.

[15] Martiniz, A.M. and R. Benavente, "The AR face database", Technical Report: 24, CVC., June, 1998.