Gender Classification Combining Facial and Hair Information

Xiao-Chen Lian and Bao-Liang Lu*

Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai 200240, China
{bllu}@sjtu.edu.cn

Abstract. Computer vision based gender classification has widespread applications. Most of the existing approaches are based on face appearance only. In this paper, we present a gender classification system that integrates face and hair features. Instead of using the whole face we extract features from eyes, nose and mouth regions with Maximum Margin Criterion (MMC), and the hair feature is represented by a fragment-based encoding. We use Support Vector Machines with probabilistic output (SVM-PO) as individual classifiers. Fuzzy integration based classifier combination mechanism is used to fusing the four different classifiers on eyes, nose, mouth and hair respectively. The experimental results show that the MMC outperforms Principal Component Analysis and Fisher Discriminant Analysis and incorporating hair feature improves gender classification performance.

1 Introduction

Gender classification is a high-level field in computer vision and is very important since its widespread applications in human-computer interaction and services that depend on it, such as demographics. There are many gender classification methods based on appearance [1–5]. Most of the existing approaches only utilize the internal facial information. We've conducted a psychological experiment: 16 participants were asked to do the gender classification task on 200 images. They were divided into two groups: Group A were fed with complete faces, and Group B were given images with only inner faces (See Fig. 1(a)). The accuracy of Group A is 100%, while the average accuracy of Group B is 92.5%. Fig. 1(b) shows images misclassified by more than 5 people of Group B. From this experiment we conclude that hair, can provide discriminative clues, especially when the face is somewhat neutral.

Following the above discussion, it is natural to expect better performance by combining face and hair information. In [6], Ueki *et al.* divided an image into face, hair, and clothing regions, and a model was learned independently for each region. The final classification of the face image was made by combining these models using a Bayesian approach.

* corresponding author

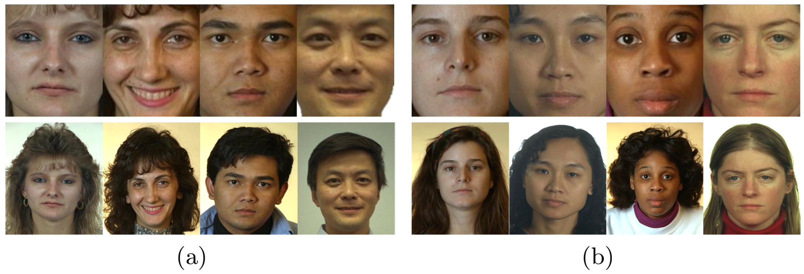


Fig. 1. Psychological experiment about gender classification: (a)experimental data; (b)images misclassified by more than 5 people of Group B

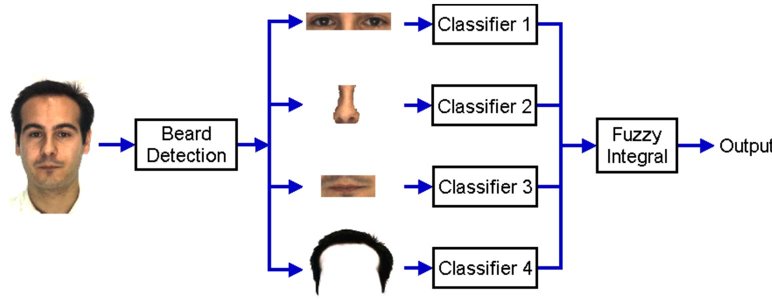


Fig. 2. The proposed gender classification system by combining four different SVMs.

In this paper, we propose a gender classification system which makes decisions by integrating face and hair information. To represent the face information, we only use the local facial features, that is, we extract the eyes, nose and mouth regions of a face and discard the rest part. Hair is hard to represent due to its large variation. Lapedriza *et al.* [7] proposed a fragment-based method which is relatively insensitive to illumination changes. We adopt this method and modify it to reduce the influence of background.

In the literature of pattern recognition, there are generally two categories of information integration approaches. One is feature combination and the other is classifier combination. We choose a fuzzy-integration-based classifier combination method. Four classifiers are trained for eyes, nose, mouth and hair features separately. Fuzzy integral is then used to combine the decisions of these classifiers into a single composite score with respect to a designated fuzzy measure. The whole process is illustrated in Fig. 2.

The remaining part of the paper is organized as follows: in section 2, face representation based on local features is introduced. In section 3, we describe the hair feature extraction method in details. Experiments and analysis are conducted in section 4, followed by conclusion and discussion in the last section.

2 Facial Feature Representation

In this section, we introduce the local-feature-based face representation method. Unlike most of the methods which use the whole face, we only consider the facial components: eyes, nose and mouth. Local features are believed very robust to the variations of facial expression, illumination, and occlusion [8]. Furthermore, psychological experiments showed that individual features (brows, eyes, nose, mouth, and chin), when seen in isolation, carried some information about gender [9].

To obtain the facial components, we adopt Active Shape Model (ASM) [10], a statistical model of the shape of the deformable object, to get the locations of eyes, nose and mouth, and then extract the rectangles centered at them respectively.

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two most widely used linear subspace learning approaches in face recognition. PCA simply performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance. It ignores the class label information and therefore is not optimal for general classification tasks. LDA, or Fisher Discriminant Analysis (FDA), is to pursue a low dimensional subspace that can best discriminate samples from different classes It requires that the within-class scatter matrix is not singular. Hence LDA is not stable facing the small sample size problem which is very common in face recognition and gender classification tasks. When applying LDA, we often have to preprocess the data with PCA.

Maximum Margin Criterion (MMC) is a recently proposed supervised feature extraction criterion [11]. Using the same notation as LDA, the goal of MMC is to maximize the criterion $J(W) = W^T(S_b - S_w)W$. The projection matrix W can be obtained by solving the following eigenvector decomposition problem:

$$(S_b - S_w)w = \lambda w \tag{1}$$

where w is the desired projection and S_b,S_w are between- and within- class scatter matrix respectively. Note that MMC does not have inverse operation and hence does not suffer from the small sample size problem.

We apply MMC to the extracted facial components. As gender classification is a 2-class problem, the resulted vector is one dimension.

3 Hair Feature Representation

Hair is represented by a fragment-based encoding. We describe the algorithm briefly. Given a collection of aligned face images C and a collection of non-face images $\overline{C}$, a representative set of image fragments called Building Blocks set (BBS) is constructed by select K most discriminative fragments among all fragments generated from regions in each image of C that contains hair. A fragments is discriminative if it appears often in C but seldom in $\overline{C}$. Some examples are shown in Fig. 3. To represent an unseen image I from BBS, for each element f_i of

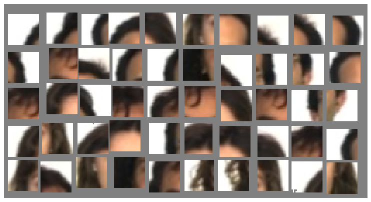
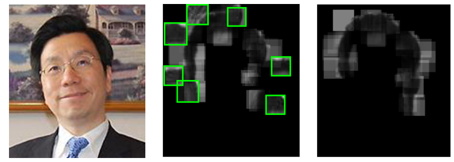


Fig. 3. Some of the building blocks.



(a) (b) (c)

Fig. 4. Illustration of the improvement.

BBS a black image B_i with the same size as I is constructed. Then f_i is located on B_i where it best matches. The similarity of two images p and f is computed by Normalized Cross Correlation (NCC). These B_i images constitute the basis and the image I is represented by linear combination of them $I \simeq \sum_{i=1}^{K} w_i B_i$. To make the combination having sense, the coefficients w_i are required non-negative. It is apparently a quadric programming aiming to minimizing the reconstruction error $(I - \sum_{i=1}^{K} w_i B_i)^2$ under the constrains $w_i \geq 0 (i = 1, \ldots, K)$. The resulted coefficient vector $W = \{w_1, \ldots, w_K\}$ encodes the hair information.

This algorithm has a problem: some fragments might be put on the background region: as shown in Fig. 4(b), there are many misplaced fragments (ones marked with green rectangles) produced by the original method with misplaced. This happens when a fragment is more similar to background than to hair in that image. Based on our experience with the implementation, the mean value of all fragments' NCC values with non-face images of $\overline{C}$ is 0.563 and the standard deviation is 0.001, and the mean value of all fragments' NCC values with face images of C is 0.818 and the standard deviation is 0.003. That means the distribution of NCC values with background and the distribution of those with hair region are separated. Therefore we discard the B_is with NCC values less than 0.750. Our modification reduces the influence of complex background (Fig. 4(c)).

4 Classifier Combination Mechanism

The concept of fuzzy integral was originally introduced by Sugeno [12] and has become increasingly popular for multi-attribute classification [13]. In this section we briefly review the concepts of fuzzy measure and fuzzy integral, and then describe how it can be applied to classifier combination.

4.1 Fuzzy Measure and Fuzzy Integral

Fuzzy measure is a extension of the classical measure where additivity property is relaxed.

Definition 1. *A fuzzy measure μ defined on $X = \{x_1, \ldots, x_n\}$ is a set function $\mu : \mathcal{P}(X) \rightarrow [0, 1]$ satisfying*
(1) $\mu(\emptyset) = 0$, $\mu(X) = 1$.

(2) $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$
$\mathcal{P}(C)$ *indicates the power set of* X*.*

Fuzzy integrals are integrals of a real function with respect to a fuzzy measure (by analogy with Lebesgue integral). There are several forms of fuzzy integrals. The one adopted here is the Choquet integral proposed by Murofushi and Sugeno [14].

Definition 2. *Let* μ *be a fuzzy measure on* X*. The discrete Choquet integral of a function* $f : X \to \mathbb{R}^+$ *with respect to* μ *is defined by*

$$C_\mu(f(x_1), \ldots, f(x_n)) \triangleq \sum_{i=1}^{n} \left(f(x_{(i)}) - f(x_{(i-1)})\right) \mu(S_{(i)}) \tag{2}$$

where $\cdot_{(i)}$ *indicates that the indices have been permuted so that* $0 \leq f(x_{(1)}) \leq \cdots \leq f(x_{(n)}) \leq 1$*. And* $S_{(i)} \triangleq \{x_{(i)} \ldots, x_{(n)}\}$*.*

4.2 Classification by Fuzzy Integral

Fuzzy integral has two advantages. One is that a number of combination mechanism are special cases of it, e.g., weighted sum, min and max rules; The other is that we can represent the importance of individual classifier and interactions (redundancy and synergy) among any subset of the classifiers using an appropriate fuzzy measure.

We describe how fuzzy integral is applied in classification tasks. Suppose $T = \{t_1, \ldots, t_m\}$ is a set of given classes. Let X be the set of classifiers and μ, a fuzzy measure defined on X, represent the contribution of each subset of X in final decision. For an unknown sample A, let $h_j(x_i)$ be the confidence of "A belongs to class t_j" given by classifier x_i. Then the global confidence of "A belongs to t_j" is given by $C_\mu(h_j(x_1), \ldots, h_j(x_n))$. We denote it by $C_\mu^j(A)$. Finally, A is given the class with highest confidence.

As mentioned above, fuzzy measure represents the contributions of classifier subsets, hence determining the fuzzy measure is a crucial step when applying fuzzy integral. For a n-classifier combination problem, we have to determine $2^n - 2$ values (fuzzy measure on empty set and X are known). These values can be learned from training data. Suppose that $(z_k, y_k), k = 1, \ldots, l$ are learning data of a 2-class problem. For simplicity, we assume that $y_k = 1$ for $k = 1, \ldots, l_1$ and $y_k = -1$ for $k = l_1 + 1, \ldots, l_2$ and $l = l_1 + l_2$. We try to identify the best fuzzy measure μ so that the squared error is minimized

$$J = \sum_{i=1}^{l_1} \left(C_\mu^1(z_i) - c_\mu^2(z_i) - 1\right)^2 + \sum_{i=l}^{l_2} \left(C_\mu^2(z_{l+i}) - c_\mu^1(z_{l+i}) - 1\right)^2 \tag{3}$$

It can be solved under quadratic program form.

Fuzzy integral requires that the output of the classifiers should be some confidence form. We choose SVM-PO with RBF kernel[15] (for the one-dimension facial features, it is equivalent to the posterior probability estimation under the assumption that the data distribution for each gender is Gaussian).

Table 1. Experimental data

Training data			Test data		
Database	female	male	Database	female	male
FERET	490	718	Postech PF01 [17]	51	54
CAS-PEAL [16]	445	595	AR	45	64
Total	935	1313	Total	96	118

Table 2. Gender classification performance of PCA, FDL and MMC

	face	eyes	nose	mouth
SVM	87.85%	85.98%	66.82%	82.71%
PCA	84.03%	84.04%	64.79%	80.75%
FDA	81.69%	83.10%	61.03%	79.34%
MMC	87.32%	85.86%	65.38%	81.69%

Table 3. Comparison of hair feature extraction methods

Methods	Ours	ICCV'05	ICPR'04
Accuracy	81.60%	79.59%	72.66%

5 Experiments

The experimental data come from frontal faces of four database with different races (See Tab. 1). We first evaluated PCA, FDA and MMC using face features (See Tab. 2). The signal-to-noise ratio of PCA was set to be 9 : 1. And the data were preprocessed by PCA for FDA. MMC utilizes the label information and therefore outperformed the other two.

Our hair feature extraction method are compared the other two methods. The first one [18] extracted the geometric features from profile of hair and the other one [6] applied gaussian mixture model on the 32×32 hair-only images. The result is shown on Tab. 3.

To investigate the relationship between fuzzy measure values and classifiers' performance, we randomly picked one fifth of training data as the validation set and computed the fuzzy measure by minimizing (3) on the validation set. Accuracy of individual classifiers on validation set is shown in Tab. 4 (Here $C_{(\cdot)}$ denotes the classifier using a particular feature). C_{eyes} performs best and the accuracy of C_{nose} is lowest. Intuitively, in the final decision the contributions of these classifiers should be proportional to their accuracy. Fuzzy measures illustrated in Tab. 5 support this intuition: the order of fuzzy measure values on individual classifiers is $\mu(\{C_{\text{eyes}}\}) > \mu(\{C_{\text{mouth}}\}) > \mu(\{C_{\text{hair}}\}) > \mu(\{C_{\text{nose}}\})$; The better a individual classifier performs, the larger fuzzy measure value on a subset containing it we obtain.

Finally, we compared gender classification performance of our method with the methods using only face information. We also compared different combination methods. Besides fuzzy integral, we chose two classifier combination mechanisms–weighted sum and product rule [19]–and one widely used feature

Table 4. Accuracy of individual classifiers on validation set

Classifier	C_{eyes}	C_{nose}	C_{mouth}	C_{hair}
Accuracy	84.25%	70.94%	78.94%	75.41%

Table 5. Fuzzy measure values of classifier subsets

Classifier subset	μ	Classifier subset	μ
$\{\emptyset\}$	0.000	$\{C_{\text{eyes}}\}$	0.691
$\{C_{\text{nose}}\}$	0.100	$\{C_{\text{mouth}}\}$	0.640
$\{C_{\text{hair}}\}$	0.148	$\{C_{\text{eyes}},C_{\text{nose}}\}$	0.691
$\{C_{\text{eyes}},C_{\text{mouth}}\}$	0.909	$\{C_{\text{eyes}},\ C_{\text{hair}}\}$	0.883
$\{C_{\text{nose}},C_{\text{mouth}}\}$	0.640	$\{C_{\text{nose}},C_{\text{hair}}\}$	0.148
$\{C_{\text{mouth}},C_{\text{hair}}\}$	0.742	$\{C_{\text{eyes}},C_{\text{nose}},C_{\text{mouth}}\}$	0.909
$\{C_{\text{eyes}},C_{\text{nose}},C_{\text{hair}}\}$	0.883	$\{C_{\text{eyes}},C_{\text{mouth}},C_{\text{hair}}\}$	1.000
$\{C_{\text{nose}},C_{\text{mouth}}C_{\text{hair}}\}$	0.742	$\{C_{\text{eyes}},C_{\text{nose}},C_{\text{mouth}},C_{\text{hair}}\}$	1.000

Table 6. Accuracy of gender classification with various methods

Fuzzy integral	Weighted sum	Product	CCA	Face
90.61%	88.73%	87.63%	74.76%	87.32%

combination method exploiting Canonical Correlation Analysis (CCA). The results are shown in Tab. 6. The last column is the performance with MMC using only face features. Fuzzy integral produced highest accuracy among these combination methods. And integrating face and hair feature through fuzzy integral improved the gender classification performance.

6 Conclusions and Future Work

One major contribution of this paper is integrating hair and face features through fuzzy-integration-based classifier combination approach. Moreover, instead of using the whole face, we extract features of facial components by MMC. The experimental results show that MMC outperforms PCA and FDA, and the gender classification benefits from incorporation of hair. A future extension of our work is to utilize clothing information. We plan to extract the profile, texture and color distribution information.

References

1. Golomb, B., Lawrence, D., Sejnowski, T.: Sexnet: A Neural Network Identifies Sex from Human Faces. Advances in Neural Information Processing Systems 3, 572–577 (1991)
2. Xia, B., Sun, H., and Lu, B.L.: Multi-view Gender Classification based on Local Gabor Binary Mapping Pattern and Support Vector Machiness. In: IJCNN 2008, PP. 3388–3395, HongKong, China (2008)

3. Lian, H.C., Lu, B.L.: Multi-View Gender Classification Using Local Binary Patterns and Support Vector Machines. International Journal of Neural System 17(6), 479-487 (2007)
4. Moghaddam, B., Yang, M.: Learning Gender with Support Faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 707–711 (2002)
5. Kim, H., Kim, D., Ghahramani, Z., Bang, S.: Appearance-based Gender Classification with Gaussian Processes. Pattern Recognition Letters 27(6), 618–626 (April 2006)
6. Ueki, K., Komatsu, H., Imaizumi, S., Kaneko, K., Sekine, N., Katto, J., Kobayashi, T.: A Method of Gender Classification by Integrating Facial, Hairstyle, and Clothing Images. In: ICPR 2004, Volume 4 (2004)
7. Lapedriza, A., Masip, D., Vitria, J.: Are External Face Features Useful for Automatic Face Classification? In: CVPR 2005, Vol. 3, pp. 151–157 (2005)
8. Su, Y., Shan., S., Chen, X., Gao, W.: Hierarchical Ensemble of Global and Local Classifiers for Face Recognition. In: ICCV 2007, pp. 1–8 (Oct. 2007)
9. BrownU, E., Perrett, D.: What Gives a Face Its Gender? Perception 22, 829–840 (1993)
10. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active Shape Models–Their Training and Application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
11. Li, H., Jiang, T., Zhang, K.: Efficient and Robust Feature Extraction by Maximum Margin Criterion. IEEE Transactions on Neural Networks 17(1), 157–165 (2006)
12. Sugeno, M., of Technology, T.I.: Theory of Fuzzy Integrals and Its Applications. PhD thesis, Tokyo Institute of Technology (1974)
13. Chang, S., Greenberg, S.: Application of Fuzzy-Integration-based Multipleinformation Aggregation in Automatic Speech Recognition. Intelligent Sensory Evaluation: Methodologies and Applications (2004)
14. Murofushi, T., Sugeno, M.: An Interpretation of Fuzzy Measures and the Choquet Integral as an Integral with Respect to a Fuzzy Measure. Fuzzy Sets and Systems 29(2), 201–227 (1989)
15. Wu, T., Lin, C., Weng, R.: Probability Estimates for Multi-Class Classification by Pairwise Coupling. The Journal of Machine Learning Research 5, 975–1005 (2004)
16. CAS-PEAL Face Database, `http://www.jdl.ac.cn/peal/index.html`
17. The POSTECH Face Database PF01, `http://nova.postech.ac.kr/special/imdb/imdb.html`
18. Yacoob, Y., Davis, L.S.: Detection and Analysis of Hair. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(7), 1164–1169 (2006)
19. Kittler, J., Hatef, M., Duin, R., Matas, J.: On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)