Neurocomputing 76 (2012) 18–27

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Gender classification by combining clothing, hair and facial component classifiers

Bing Li [a], Xiao-Chen Lian [a], Bao-Liang Lu [a,b,*]

[a] Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

[b] MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

ARTICLE INFO

Available online 22 August 2011

Keywords:
Gender classification
Facial components
Clothing feature
Hair feature
Classifier combination
Local binary pattern
Support vector machine

ABSTRACT

In this paper, we propose a novel gender classification framework, which utilizes not only facial features, but also external information, i.e. hair and clothing. Instead of using the whole face, we consider five facial components: forehead, eyes, nose, mouth and chin. We also design feature extraction methods for hair and clothing; these features have seldom been used in previous work because of their large variability. For each type of feature, we train a single support vector machine classifier with probabilistic output. The outputs of these classifiers are combined using various strategies, namely fuzzy integral, maximal, sum, voting, and product rule. The major contributions of this paper are (1) investigating the gender discriminative ability of clothing information; (2) using facial components instead of the whole face to obtain higher robustness for occlusions and noise; (3) exploiting hair and clothing information to facilitate gender classification. Experimental results show that our proposed framework improves classification accuracy, even when images contain occlusions, noise, and illumination changes.

1. Introduction

Gender classification using facial images is widely used in applications like human–computer interface, demographics, and customer-oriented advertising. The most common pipeline for gender classification is feature extraction (including preprocessing) followed by classifiers trained on the features. In recent years, based on this pipeline, many approaches have been proposed [1–9], which have achieved promising accuracy on various data sets.

However, practical applications of these approaches still suffer from two unsolved problems. The first is the neutral face problem. As the name suggests, neutral faces are those whose appearance conveys less discriminative information. From the point of view of pattern recognition, neutral faces are the points which always lie on the boundary between male and female regions in the feature space, no matter what kind of feature representation is adopted. The other problem is the presence of face occlusions, e.g. masks and sunglasses, and noise and illumination changes. In feature space, they cause a big deviation from "contaminated" face space to normal face space.

Humans do not classify gender based solely on facial information; they also use other information such as hair, clothing, voice and gait when encountering difficulties. The psychological experiment in our previous work showed that hair provides discriminative clues to human [10], especially when the face is neutral. To prove the effectiveness of clothing, we conduct a similar experiment: First, 548 upper body clothing images without faces (half male and half female) as shown in Fig. 1 are collected. Then three human subjects are asked to classify the gender of these images. They get 96.3% classification accuracy on color images, and 93.5% on gray images on average. From this experiment, we conclude that clothing can serve as a useful clue for humans in gender classification tasks.

To overcome the occlusion problem, a strategy is to extract local features instead of global features. This strategy is the foundation of the bag-of-words framework for object classification [11] to deal with the large intra-class variations of objects. A drawback of using local features is its ignorance of some global information, especially spatial constraints. Therefore, determining the granularity of the local features is a crucial issue: the scale of features cannot be on the pixel level which bag-of-words methods use, nor can it be too large since that would lose the robustness against variations. In [12], psychological experiments showed that individual facial components (brows, eyes, nose, mouth and chin), when seen in isolation, carried much information about gender. Therefore facial component level is a good choice.

Based on the discussion above, we propose a novel gender classification system, which utilizes hair and clothing information and extracts facial features from facial components instead of the

* Corresponding author at: Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.
E-mail addresses: bllu@sjtu.edu.cn, blu@cs.sjtu.edu.cn (B.-L. Lu).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.01.028

Fig. 1. Upper body clothing images without faces.

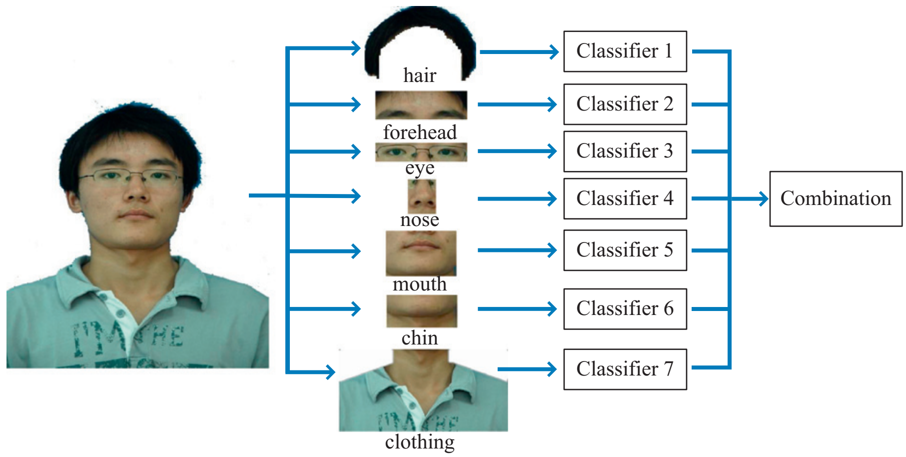


Fig. 2. Flowchart of our gender classification framework.

whole face. The flowchart of the framework is shown in Fig. 2. First, in the feature extracting stage, the facial region is divided into five facial components, namely forehead, eyes, nose, mouth, and chin. The feature representations for hair and clothing are also designed. Then in the classifier training stage, seven classifiers are trained correspondingly. Here we use support vector machines (SVMs) with probabilistic outputs [13]. Finally, five classifier combination strategies (fuzzy integral [14], sum, maximal, voting and product rule) are used to integrate the results of the seven classifiers into a single composite score for gender prediction.

In our experiments, FERET [15], AR [16] and a data set collected by us (denoted by BCMI[1]) are used. Three multi-feature combination experiments are done: (a) combination of five facial components (O); (b) combination of clothing, hair and the whole face (C.H.F); and (c) combination of clothing, hair and five facial components (C.H.O). Comparisons of the results show that the facial components combination gives better accuracy than the whole face that the clothing, hair and the whole face combination performs better than any single feature, and that the clothing, hair and five facial components combination achieves the best accuracy. Experiments also indicate that our proposed multi-feature combination framework performs much better than Ueki's method [17], which is another work using clothing and hair information to facilitate gender classification.

To demonstrate the robustness of our framework, we carry out experiments on two kinds of test images. One, based on the BCMI data set, has artificial occlusions and Gaussian white noise. The other, based on the AR data set. has natural occlusions (glasses and scarves) and illumination changes. One experiment is to compare the gender classification accuracy of the facial component classifiers combination with that of the whole face classifier. The other is to compare the accuracy with clothing and hair features to that without. These experiments show that our framework is robust against occlusions and noise.

The remaining part of this paper is organized as follows. In Section 2, we introduce related work on gender classification. In Section 3, we introduce clothing, hair and facial component representation methods. In Section 4, we describe pattern classifiers and five classifier combination strategies. In Section 5, we evaluate experimental results. Finally, we give our conclusions and discuss future work.

2. Related work

Various gender classification methods are reported in the literature. These methods can be divided to two main categories.

The first category is appearance-based approaches. Some early work used single- or multiple-layer neural networks with image pixels as input, e.g. [1,18–20,3]. Sun et al. [21] computed Principle Component Analysis (PCA) features from faces while Jain and Huang [22] used independent component analysis (ICA). Moghaddam and Yang [4] demonstrated that SVM work much better in gender classification than RBF networks and nearest neighbor. Lian and Lu [5,23] used min-max modular SVM with image pixels as input. They also experimented with Local Binary Pattern (LBP) as face features. Xia et al. [8] adopted Local Gabor Binary Papping Pattern to extract face feature. Saatci and Town [24] build the Active Appearance Model (AAM) for faces and developed a cascaded structure of SVMs for gender classification. Kim et al. [6] applied Gaussian process method, as it could automatically determine the hyper-parameters. Adaboost framework is also utilized in [25,26,7] and achieved promising accuracy. Erno and Roope [9] have done a comprehensive evaluation of state-of-the-art appearance-based approaches combined with automatic real-time face detection and manual face normalization.

The other category is based on geometrical features. Brunelli and Poggio [2] extracted 16 geometric features from faces, such as eyebrow thickness and pupil to eyebrow separation, as input to the HyperBF network to learn the differences between the two genders. Burton et al. [27] extracted point-to-point distances from 73 points on face images and used discriminant analysis as a classifier.

Although having achieved good performances in gender classification on various data sets, the approaches mentioned above

[1] The database is named after the abbreviation of our research center.

only used facial region and discarded other information such as clothing and hair because of their variability. To the best of our knowledge, only a few works have discussed using multiple sources of information in gender classification. Ueki et al. [17] extracted clothing, hair and facial information using principal component analysis (PCA), trained Gaussian mixture models (GMM) for two clothing styles, and used a Bayesian approach to combine these models for gender classification. However, GMM models are susceptible to over-fitting. Our previous work involved hair information in gender classification. Ji et al. [28] constructed a geometric hair model (GHM) to extract hair features and used local binary pattern (LBP) to extract facial features, then a support vector machine to classify gender. Classification accuracy was improved by combining hair and facial information. In [10], a fragment-based hair representation was used, and hair and facial information were combined by fuzzy integral. It highly improved classification accuracy.

3. Feature extraction of clothing, hair, and facial components

3.1. Local binary pattern (LBP) descriptor

Local binary patterns were first used by Ahonen et al. [29] for face recognition and by Lian and Lu [5] for gender classification. It is a very simple but efficient algorithm for local texture information extraction and it is also stable under illumination changes and rotation. We use this method for clothing and facial feature extraction.

The original local binary pattern operator labels the pixels of an image by thresholding the 3×3 or more neighborhood of each pixel with the center value as illustrated in Fig. 3(a) and considering the result as a binary number:

$$S(f_p - f_c) = \begin{cases} 1, & f_p - f_c \geq 0 \\ 0, & f_p - f_c < 0 \end{cases} \quad (1)$$

where f_c is the value of the center and $f_p (p = 0, 1, \ldots, 7)$ is the value of the neighborhood of f_c. The LBP operator value at the center pixel is

$$LBP(f_c) = \sum_{p=0}^{7} S(f_p - f_c) 2^p. \quad (2)$$

A 3×3 neighborhood example is shown in Fig. 3(b). The final LBP operator binary value is 11010011, which is 211 in decimal. In 3×3 neighborhood, LBP operator ranges from 0 to 255. Mapping the LBP operator value in every pixel to a gray value, we get an LBP texture image. The LBP texture images of Fig. 1 are shown in Fig. 3(c).

A uniform pattern which is an extension to the original LBP operator is used in this paper. A local binary pattern operator is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000 and 11000011 are uniform LBP codes, but 10101111 is not uniform. There are 58 labels for uniform patterns and one label for any non-uniform pattern, so 59 labels for the 3×3 neighborhood are used overall. The image is divided into m regions $R_0, R_1, \ldots, R_{m-1}$ and the spatially enhanced histogram is defined as:

$$H_{j,i} = \sum_{x,y} I\{f_{ul}(x,y) = i\} I\{(x,y) \in R_j\}, \quad i = 0, 1, \ldots, n-1; \; j = 0, 1, \ldots, m-1, \quad (3)$$

where

$$I\{A\} = \begin{cases} 1, & \text{if A is true} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where n is the number of uniform patterns, n is 59 in a 3×3 neighborhood, and $f_{ul}(x,y)$ is the uniform pattern value in position (x,y). Combining these feature vector of sub-regions, the final vector feature is obtained:

$$V = \{H_{0,0}, \ldots, H_{0,n-1}, H_{1,0}, \ldots, H_{1,n-1}, \ldots, H_{m-1,0}, \ldots, H_{m-1,n-1}\}. \quad (5)$$

LBP features can be described on three levels: an LBP uniform operator describes the pattern on the pixel level, the labels summed over a sub-region give information on the regional level, and the combination of sub-region histograms gives a description on the global level. LBP features can efficiently represent image texture information at all levels, so the technique is widely used in many applications.

3.2. Clothing feature extraction

Extracting effective features from clothing regions is challenging since faces have a relatively constant structure (e.g. almost all non-occluded faces have brows, eyes, nose and mouth), while clothing is much more complex and can have a large variety of styles in texture, shapes, and colors. The same clothes may have different visual effects when worn by different persons or even the same person moving differently. Clothing information is usually discarded in existing gender classification systems due to its large variation. There is little research on extracting clothing features. In spite of the differences in individual clothing, we observe that clothing from the same gender does share some common features. For instance, male clothing usually has a simple style and color, with a shirt with a collar and fewer ornaments, while females may have complex styles such as open collar, lots of ornaments, and multiple colors. Taking the four clothing images in Fig. 1 as an example, we can easily see that the former two are male and the latter two are female. The two males are

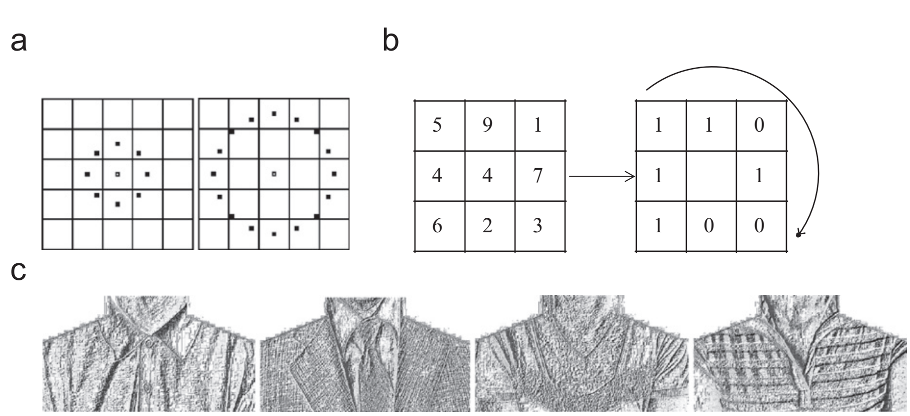


Fig. 3. Local binary pattern and clothing texture images: (a) LBP circular, (b) a LBP operator, (c) clothing texture images after LBP operation (original images shown in Fig. 1).

wearing shirts with simple color and the two females are wearing clothes with complex styles, open collar and multiple colors.

Motivated by the above observation, we use a color histogram and the local binary pattern algorithm for clothing feature extraction in this paper. Color histogram involves color information of clothing and local binary pattern [30] involves texture information. Both methods are invariant to rotation and illumination changes.

3.3. Hair feature extraction

Traditional face recognition and gender classification systems usually discard hair information because of its large variation in geometry, colors and texture. However, this information is highly useful in gender classification because of the obvious difference in hair length and style between male and female. Based on our previous work [10], we adopt Lapedriza's method [31] for hair feature extraction. In this method, hair zones are selected to form a building block set, which is used to represent the unseen image as a set of puzzle pieces. Then the unseen image is reconstructed with the most similar fragments in the building block set. The hair feature is encoded by a weight vector which represents the importance of every fragment.

3.4. Facial component feature extraction

Facial information is important in representing human individuality and gender. In the literature, feature extraction is usually applied on the whole face, and the results are called global features. However, the quality of global features greatly depends on face alignment and occlusions, while local features are believed robust to the variation of facial expression, illumination, and occlusions [32]. Furthermore, psychological experiments show that sub-face individual features (brows, eyes, nose, mouth, and chin), when seen in isolation, carry much gender information [12]. In this paper, we consider using the features of forehead, eyes, nose, mouth, and chin components, separately. Since the local binary pattern technique has been successfully applied to face feature extraction for face recognition [29] and gender classification [5], here we use a local binary pattern algorithm for extracting the whole face and facial component features. We detect facial regions using the Adaboost face detection algorithm [33] and locate facial components using an active shape model (ASM) [34] automatically, then the five facial components (forehead, eyes, nose, mouth and chin) are cropped to fixed window sizes. The details of image preprocessing and cropping will be described in Section 5.

4. Pattern classifier and classifier combination strategies

Support vector machines (SVMs) are the most commonly used pattern classifier in gender classification [4]. We use SVMs with probabilistic output [13]. Suppose we have training data $\mathcal{D}^t=\{(x_i^t,d_i^t)\}_{i=1}^N$ for a component, where $t\in$ {clothing, hair, forehead, eyes, nose, mouth, and chin}, x_i^t is the component feature extracted from the i-th image, and $d_i^t\in\{-1,1\}$ are gender labels (1 for male and -1 for female). We train a component classifier g^t using $\mathcal{D}^t$. The output of g^t is in the range $[-1,1]$ with the sign indicating gender and absolute value indicating the confidence.

Then a classifier integration mechanism combines the results of these classifiers. In the following, we first introduce four simple but widely used integration methods—maximal, sum, voting, and product rule, then we introduce fuzzy integral, which has larger expressive ability and in special cases is equivalent to those four methods.

4.1. Simple combination strategies

The advantage of maximal, sum, voting and product rules is their simplicity: they require no training. Given the outputs of all the classifiers, the maximal (sum, product rule) method is to compare the maximum (sum, product) of all the probability confidences that a sample belongs to male and female in all the classifiers; and the voting method obtains the prediction result by voting from all the classifiers.

4.2. Fuzzy-integration-based classifier combination

The defect of these simple strategies is that they rely on the unreasonable assumption that all the component classifiers are mutually independent. This assumption is usually inconsistent with the real situation.

To eliminate such assumptions, Sugeno [14] introduced the concept of the fuzzy integral, which has become increasingly popular for multi-attribute classification. A fuzzy integral actually is the integration of real functions with fuzzy measures. Fuzzy measures are an extension of classical measures. A general fuzzy measure is defined as follows:

Definition 1. A fuzzy measure μ defined on $X=\{x_1,x_2,\dots,x_n\}$ is a set function μ: $P(X)\to[0,1]$ ($P(X)$ indicates the power set of X) satisfying:

(1) $\mu(\emptyset)=0, \mu(X)=1$,
(2) $A\subseteq B\Rightarrow\mu(A)\le\mu(B)$.

The fuzzy measure we adopt in this paper is the Choquet integral [35].

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

$$C_\mu(f(x_1),\dots,f(x_n))\triangleq\sum_{i=1}^{n}(f(x_i)-f(x_{i-1}))\mu(S_i), \tag{6}$$

where i indicates that the indices have been permuted so that $0=f(x_0)\le f(x_1)\le\cdots\le f(x_n)\le 1$ and $S(i)\triangleq\{x_i,\dots,x_n\}$.

Fuzzy integrals have two advantages. First, with a properly designed fuzzy measure, a fuzzy integral can simulate any one of the four methods described in Section 4.1. For example, with an evenly distributed measure on all subsets, the fuzzy integral is actually the "maximal" operator. Second, we can represent the importance of individual classifiers and interactions (redundancy and synergy) among any subset of the classifiers using an appropriate fuzzy measure.

The expressive ability lies in the fact that fuzzy measures can be set freely, as long as they do not violate the two constraints in Definition 2. We can compute the fuzzy measure μ from training data, by minimizing error J:

$$J=\sum_{i=1}^{m}(C_\mu^1(x_i)-C_\mu^{-1}(x_i)-1)^2+\sum_{j=1}^{n}(C_\mu^{-1}(y_j)-C_\mu^1(y_j)-1)^2, \tag{7}$$

where $\{x_0,x_1,\dots,x_m\}$ are m samples that belong to class 1, $\{y_0,y_1,\dots,y_n\}$ are n samples that belong to class -1, and $C_\mu^k(z)(k=-1,1)$ is the global confidence that z belongs to class k by classifier combination, which is given by

$$C_\mu^k(z)=C_\mu(h_k(g_1),\dots,h_k(g_c)), \tag{8}$$

where C_μ is defined as Definition 2, μ is the fuzzy measure, $h_k(g_t)$ is the confidence of sample z belonging to class k in classifier g_t, $t\in\{1,2,\cdots,c\}$, and c is the number of classifiers to be combined. This is actually a quadratic optimization problem.

5. Experiments

Two data sets are used in our experiment. One is FERET [15] and the other is BCMI collected by us. The images in the BCMI data set contain clear human frontal upper clothing, facial and hair regions. Some examples are shown in Fig. 4. All of the images are of eastern Asians, of different ages and occupations. In the FERET data set, 782 images with clear clothing regions are chosen. Four hundred and fifty four of them are chosen randomly as training data, and the rest as test data. The BCMI data set contains 2190 images in total. One thousand six hundred and forty two images are chosen randomly as training data, and the rest as test data. In both data sets, the ratio of training to test data is about 3:1, and both the training and test data have equal numbers in the two genders, as shown in Table 1. In the training stage, four fifths of the training data are picked randomly for training SVM models and the remaining fifth is used for computing the fuzzy measure by minimizing Eq. (7) on it.

All of these experiments are repeated ten times with different random partitions of the data. The final results are reported as the mean and standard deviation of the results from individual runs. The parameters of SVMs are determined by five-fold cross validation, and three kernels, namely linear, polynomial, and RBF, are used in our experiments. The code for SVMs is from LibSVM [36]. The experiments are performed on a computer with 8G RAM and a 2.83 GHz CPU.

5.1. Image preprocessing

The flowchart of preprocessing is shown in Fig. 5(a). First, the background is removed and some important points on brows, eyes, nose, and mouth are located. In the first step, we apply the Sobel edge detection algorithm to the input image to obtain the edge contour. According to the edge contour the background is removed. To locate important points, the facial region is first detected by the Adaboost face detector [33], and then the locations of brows, eyes, nose, mouth, and chin are obtained from the Active Shape Model (ASM) [34] previously trained. After that, we rotate the image so that eyes are on the same horizontal line, and resize it to make the size of the facial region 150×130 pixels. Finally facial components and clothing regions are cropped according to the fixed window sizes shown in Fig. 5(b).

5.2. Clothing information

We extract clothing information using the LBP algorithm. The image is divided into 7×7 sub-regions with 3×3 neighborhood, yielding a feature vector with dimension of $7 \times 7 \times 59 = 2891$. The average extraction time for an image is about 75 ms, as shown in Table 3. Results of gender classification on clothing features extracted by gray, color histogram, and LBP algorithm are shown in Table 2. The average accuracies of LBP feature are 72.3% and 79.5% on the two data sets.

5.3. Hair and facial information

To extract hair information, 600 hair fragments with size of 10×10 pixels extracted from images in training data are selected as the Building Blocks set, and then hair features are calculated using this set. LBP is applied to extract both global and local face features. The number of LBP bins, feature vector dimensions and feature extraction time for hair, clothing, face and facial components are shown in Table 3. We can see that hair feature extraction takes the longest time due to the high computation complexity of matrix factorization.

5.4. Single classifier

For each component, an SVM classifier with linear, polynomial, and RBF kernels is trained. The total training time for each SVM and average test time of a sample are shown in Table 4. Each SVM gender model needs only be trained offline once, then the gender

Table 1
Description of training and test data sets from FERET and BCMI data set.

Database	No. male/no. female	Total	No. training	No. test
FERET	341/341	682	227×2	114×2
BCMI	1095/1095	2190	821×2	274×2

Fig. 4. Sample examples of BCMI data set.

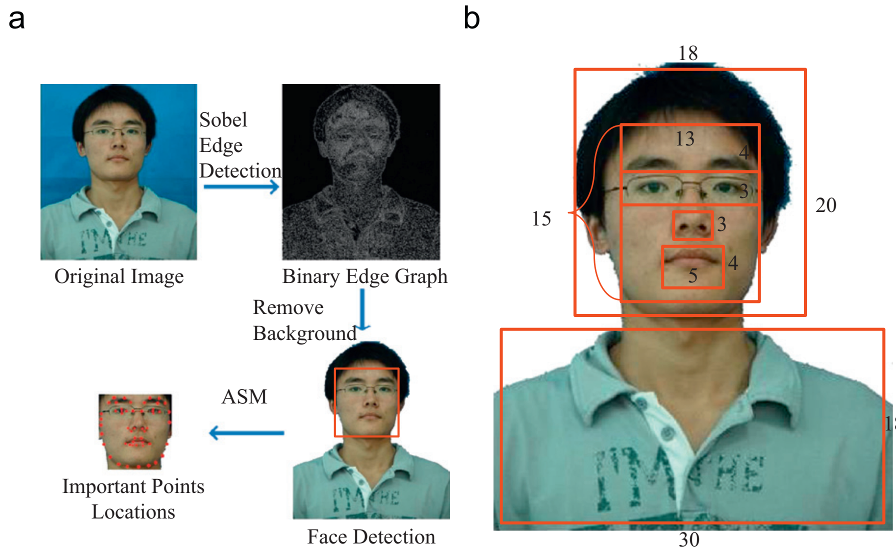


Fig. 5. Image preprocessing: (a) the flowchart of image preprocessing and (b) window sizes for cropping.

Table 2
Accuracy (%) of gender classification based on clothing features.

Kernel	FERET			BCMI		
	Gray-His	Color-His	LBP	Gray-His	Color-His	LBP
Linear	61.6 ± 1.6	64.3 ± 1.3	$\mathbf{73.0 \pm 0.8}$	66.0 ± 2.0	70.2 ± 1.2	$\mathbf{76.3 \pm 1.4}$
Poly	63.7 ± 2.3	65.8 ± 1.1	$\mathbf{71.8 \pm 2.6}$	67.7 ± 1.0	72.5 ± 0.8	$\mathbf{80.3 \pm 1.2}$
RBF	63.6 ± 2.1	66.2 ± 1.1	$\mathbf{72.2 \pm 2.8}$	68.3 ± 1.4	72.2 ± 0.8	$\mathbf{81.8 \pm 1.3}$
Average	63.0 ± 2.0	65.4 ± 1.2	$\mathbf{72.3 \pm 2.1}$	67.3 ± 1.5	71.7 ± 0.9	$\mathbf{79.5 \pm 1.3}$

Table 3
LBP bins, feature vector dimensions (Dim) and extraction time (ms) of hair, clothing, face and facial components.

Parameter	Clothing	Hair	Face	Forehead	Eyes	Nose	Mouth	Chin
Bins	7×7	–	7×7	3×7	2×7	3×2	3×5	2×5
Dim	2891	600	2891	885	590	352	1180	885
Time	74.4	1540	25.0	7.8	4.7	1.6	9.4	6.3

of a test sample can be predicted in a few milliseconds. This means that our framework can be applied to a real-time gender classification system. Because these SVM models are independent of each other, the framework could be easily implemented in parallel, which would further reduce the computation time.

The performance of each classifier is shown in Table 5. We can see that the classifier of the whole face feature gets the highest accuracy, which is 88.5% and 91.9% on average on the two data sets. In both data sets, among facial components classifiers, forehead and chin perform best, while nose does worst. This indicates that forehead and chin regions provide the most discriminative information for gender, while nose is comparatively less useful. In addition, the gender classifier using hair gets accuracy of 79.7% and 77.7%, which indicates our hair feature representation is effective.

5.5. Classifier combination

We consider three classifier combinations: (1) combination of clothing (C), hair (H), and the whole face (F); (2) combination of five facial components (O); and (3) combination of clothing (C), hair (H), and five facial components (O). They are denoted by (C.H.F), (O), and (C.H.O) respectively. We adopted five different combination mechanisms: sum, maximal, voting, product rule and fuzzy integral. The results are shown in Table 6. On FERET data set, fuzzy integral performs best in all cases. On BCMI data set, product rule performs better than fuzzy integral in combinations (C.H.F) and (O), while fuzzy integral outperforms other cases. By comparing results in Tables 5 and 6, we can see that combining classifiers improves the performance. An interesting result is that the accuracies obtained by combination of facial components, are higher than those obtained by the face classifier which uses global features extracted on the whole face. In Section 5.6, we will demonstrate that with occlusions and noise, the superiority of (O) to the whole face classifier is more significant.

Finally, we compare our method with Ueki's [17]. The classifier combination method we use here is fuzzy integral. The results are shown in Table 7. In Ueki's work, PCA is used to extract features and the classifier is GMM. From this table, we can see that for hair and face classifiers, our method performs better (clothing result for Ueki's is not shown since in their work the corresponding classifier is not trained for classifying gender). When combining individual classifiers, our method performs much better.

5.6. Robustness of multi-feature combination

We first evaluate the robustness against occlusions and noise gained by using facial components instead of the whole face. We carry out two experiments on artificial and natural low quality images.

Table 4
Training time (h) of SVM classifiers and average testing time (ms) of a sample image.

Type	Kernel	Clothing	Hair	Forehead	Eyes	Nose	Mouth	Chin	Face
Training (h)	Linear	2.02	1.97	0.30	0.32	0.28	0.62	0.77	2.38
	Poly	2.72	1.98	0.48	0.33	0.25	0.68	0.77	3.08
	RBF	5.20	2.53	0.73	0.58	0.32	1.17	1.33	5.87
	Average	3.31	2.16	0.50	0.41	0.28	0.82	0.96	3.78
Test (ms)	Linear	13.02	2.82	3.55	2.87	2.31	5.49	4.92	11.49
	Poly	15.01	3.34	3.60	2.81	2.52	6.34	5.42	13.37
	RBF	22.52	3.26	4.67	3.29	2.60	8.76	6.45	24.62
	Average	16.85	3.14	3.94	2.99	2.48	6.86	5.60	16.49

Table 5
Gender classification accuracy (%) of SVM classifiers on single feature.

Database	Kernel	Hair	Forehead	Eyes	Nose	Mouth	Chin	Face
FERET	Linear	78.5 ± 1.1	82.9 ± 1.4	73.8 ± 0.7	65.5 ± 2.0	82.8 ± 0.8	82.6 ± 1.1	$\mathbf{88.6 \pm 1.1}$
	Poly	80.1 ± 1.3	83.3 ± 1.0	75.1 ± 0.6	63.2 ± 2.7	83.5 ± 0.6	83.5 ± 0.8	$\mathbf{88.3 \pm 0.7}$
	RBF	80.6 ± 1.5	82.9 ± 1.0	76.2 ± 2.0	64.9 ± 1.4	82.6 ± 1.7	84.7 ± 1.6	$\mathbf{88.6 \pm 1.1}$
	Average	79.7 ± 1.3	82.6 ± 1.1	75.1 ± 1.1	64.5 ± 2.0	82.9 ± 1.1	83.6 ± 1.2	$\mathbf{88.5 \pm 1.0}$
BCMI	Linear	76.8 ± 1.4	86.1 ± 1.5	81.3 ± 2.2	66.4 ± 1.8	82.2 ± 1.0	84.3 ± 1.1	$\mathbf{91.6 \pm 0.8}$
	Poly	77.8 ± 1.1	86.9 ± 1.0	82.6 ± 2.0	68.7 ± 1.6	83.6 ± 1.1	85.7 ± 1.1	$\mathbf{91.9 \pm 0.9}$
	RBF	78.4 ± 0.8	87.5 ± 0.9	82.9 ± 2.6	68.3 ± 1.5	83.5 ± 1.2	85.9 ± 1.4	$\mathbf{92.2 \pm 1.1}$
	Average	77.7 ± 1.1	86.8 ± 1.1	82.3 ± 2.3	67.8 ± 1.6	83.1 ± 1.1	85.3 ± 1.2	$\mathbf{91.9 \pm 0.9}$

Table 6
Gender classification accuracy (%) of classifiers combinations: (C.H.F), (O) and (C.H.O) using sum, maximal, voting, product rule and fuzzy integral on FERET and BCMI.

Method	Database	Kernel	Sum	Maximal	Voting	Product	Fuzzy
(C.H.F)	FERET	Linear	90.2 ± 1.5	89.9 ± 2.3	89.6 ± 0.2	90.8 ± 1.8	$\mathbf{92.9 \pm 0.2}$
		Poly	91.4 ± 1.3	90.8 ± 1.2	89.9 ± 0.4	90.9 ± 1.3	$\mathbf{92.7 \pm 0.3}$
		RBF	90.8 ± 2.3	90.2 ± 2.0	90.1 ± 2.1	91.2 ± 2.0	$\mathbf{93.4 \pm 0.9}$
		Average	90.8 ± 1.7	90.3 ± 1.8	89.9 ± 0.9	91.0 ± 1.7	$\mathbf{92.7 \pm 0.5}$
	BCMI	Linear	92.9 ± 0.7	93.3 ± 0.8	91.6 ± 1.2	$\mathbf{93.6 \pm 0.8}$	93.2 ± 0.5
		Poly	93.9 ± 0.8	93.9 ± 0.9	93.4 ± 1.2	$\mathbf{94.5 \pm 0.7}$	93.4 ± 0.6
		RBF	93.6 ± 1.0	94.3 ± 1.1	93.6 ± 1.2	$\mathbf{94.4 \pm 1.0}$	93.6 ± 1.0
		Average	93.4 ± 0.8	93.8 ± 0.9	92.9 ± 1.2	$\mathbf{94.2 \pm 0.8}$	93.4 ± 0.6
(O)	FERET	Linear	89.5 ± 1.2	90.8 ± 2.1	87.3 ± 1.6	91.2 ± 2.0	$\mathbf{91.7 \pm 0.9}$
		Poly	89.5 ± 1.8	89.9 ± 2.3	86.8 ± 1.9	90.4 ± 2.0	$\mathbf{93.0 \pm 0.4}$
		RBF	90.8 ± 2.8	91.2 ± 2.9	89.0 ± 2.9	91.2 ± 2.7	$\mathbf{93.4 \pm 0.9}$
		Average	89.9 ± 1.9	91.2 ± 2.4	87.7 ± 2.1	90.9 ± 2.2	$\mathbf{92.7 \pm 0.7}$
	BCMI	Linear	92.3 ± 1.5	91.3 ± 1.6	91.2 ± 1.7	92.7 ± 1.5	$\mathbf{92.9 \pm 1.2}$
		Poly	93.2 ± 1.2	92.8 ± 1.4	91.9 ± 1.4	$\mathbf{93.8 \pm 1.1}$	93.4 ± 1.0
		RBF	93.4 ± 1.1	93.2 ± 1.4	92.5 ± 1.6	$\mathbf{94.0 \pm 1.1}$	93.7 ± 1.0
		Average	93.0 ± 1.3	92.5 ± 1.5	91.9 ± 1.6	$\mathbf{93.5 \pm 1.2}$	93.3 ± 1.1
(C.H.O)	FERET	Linear	91.7 ± 2.1	90.8 ± 2.0	91.2 ± 2.3	91.2 ± 1.8	$\mathbf{94.9 \pm 0.6}$
		Poly	90.8 ± 1.5	90.4 ± 1.8	90.8 ± 2.9	91.2 ± 1.6	$\mathbf{95.8 \pm 0.4}$
		RBF	91.7 ± 2.7	93.4 ± 3.3	91.7 ± 4.0	93.4 ± 3.6	$\mathbf{94.7 \pm 0.5}$
		Average	91.4 ± 2.1	91.5 ± 2.4	91.2 ± 3.1	92.0 ± 2.3	$\mathbf{95.1 \pm 0.5}$
	BCMI	Linear	93.7 ± 0.9	91.8 ± 1.3	91.0 ± 1.4	93.5 ± 1.0	$\mathbf{94.5 \pm 0.9}$
		Poly	94.3 ± 1.0	93.1 ± 1.4	92.7 ± 0.9	94.6 ± 1.0	$\mathbf{95.3 \pm 0.6}$
		RBF	94.6 ± 1.0	94.2 ± 1.2	93.1 ± 1.0	94.9 ± 0.9	$\mathbf{95.3 \pm 0.8}$
		Average	94.2 ± 1.0	93.0 ± 1.3	92.3 ± 1.1	94.3 ± 0.9	$\mathbf{95.0 \pm 0.8}$

Table 7
Gender classification accuracy (%) comparison with Ueki's method.

Database	Method	Clothing	Hair	Face	(C.H.F)	(O)	(C.H.O)
FERET	Ours	72.3 ± 2.1	79.7 ± 1.3	88.5 ± 1.0	92.7 ± 0.5	92.7 ± 0.7	$\mathbf{95.1 \pm 0.5}$
	Ueki's	–	58.5 ± 3.9	78.5 ± 1.8	$\mathbf{82.0 \pm 0.5}$	–	–
BCMI	Ours	79.5 ± 1.3	77.7 ± 0.8	91.9 ± 1.1	94.2 ± 1.0	93.5 ± 1.1	$\mathbf{95.0 \pm 0.8}$
	Ueki's	–	73.3 ± 1.7	79.2 ± 5.6	$\mathbf{87.6 \pm 2.0}$	–	–

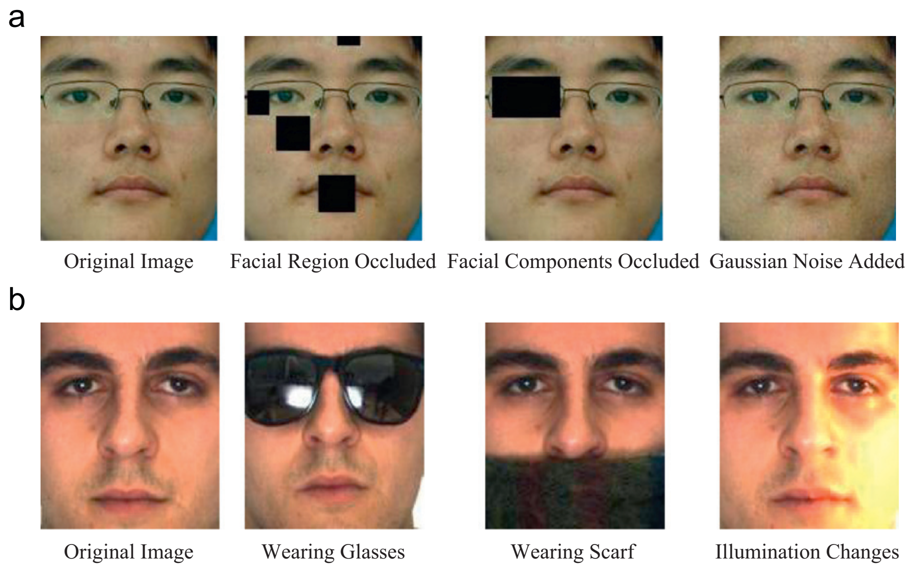


Fig. 6. Examples of experiment data: (a) images with artificial occlusions and Gaussian white noise in BCMI and (b) images with natural occlusions and illumination changes in AR.

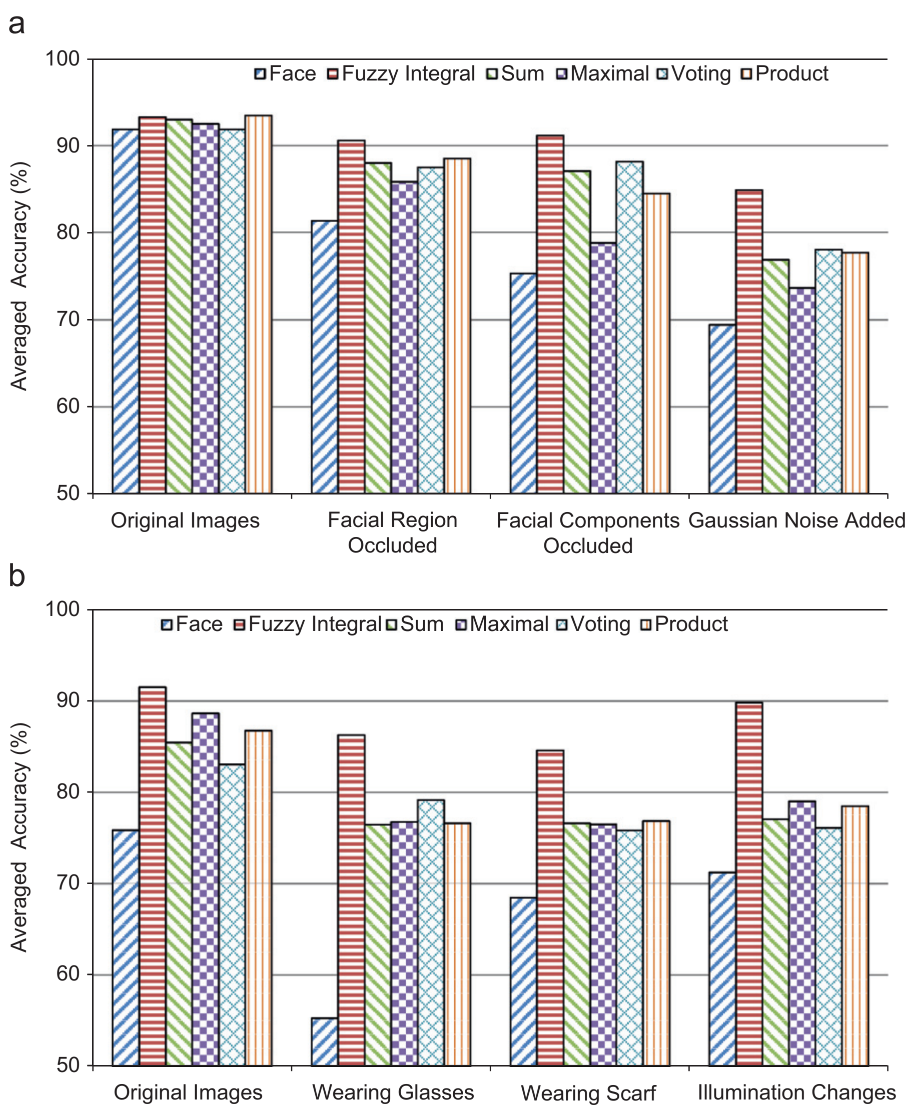


Fig. 7. Comparison of (O) with face feature classifier on original and three kinds of contaminated images: (a) experiments on images with artificial occlusions and Gaussian white noise and (b) experiments on images with natural occlusions and illumination changes.

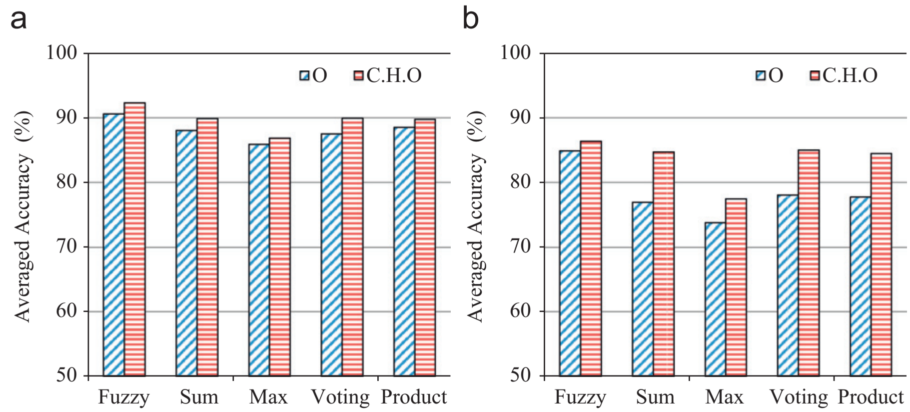


Fig. 8. Comparison of gender classification accuracy (%) of (C.H.O) with that of (O) on images with random occlusions and Gaussian noise on clothing, hair and facial regions: (a) random occlusions and (b) Gaussian noise.

The first experiment is on the BCMI data set. We keep the original training images, but artificially process each test image in one of the following three ways (as shown in Fig. 6(a)):

(a) Occlude facial regions randomly: four sub-regions of face are randomly selected to be occluded with black rectangles with random sizes (ranging from 5 $\times$ 5 pixels to 20 $\times$ 20 pixels).
(b) Occlude facial components randomly: one of the following: left eye, right eye, nose, and mouth component regions is randomly selected to be occluded with a rectangle generated in the same way as (a).
(c) Gaussian noise: add Gaussian white noise (with zero mean and 0.01 variance) to sample images.

The second experiment is on the AR face data set [16]. We keep the original training images from the FERET data set, and use four kinds of test images from the AR data set: original images, images wearing glasses, wearing a scarf and illumination changes, as the examples shown in Fig. 6(b). In this test set, there are 76 male and 59 female samples.

Fig. 7 compares the performance of the face classifier with those of facial component classifiers combination (O) strategies: fuzzy integral, maximal, sum, voting and product rule, on artificial (Fig. 7(a)) and natural (Fig. 7(b)) test data. Here we report accuracies averaged over three SVM kernels (linear, polynomial and RBF). All classifiers suffer from occlusions, noise and illumination changes. However, all the facial component classifiers' combinations outperform the whole face classifier, which proves that using facial components strengthens robustness against occlusions, noise and illumination changes. Moreover we can see that fuzzy integral is the most stable combination strategy among the five combination methods.

To evaluate the robustness gained by combining hair, clothing, and face, we conduct a similar experiment. Besides facial regions, we also add random occlusions and noise on clothing and hair regions. We compare the gender classification accuracy of the clothing, hair and facial components combination (C.H.O) with that of the facial components combination (O). The experimental results (shown in Fig. 8) demonstrate that (C.H.O) performs better than (O) when test images are contaminated. Therefore, we can conclude that our proposed gender classification framework improves gender classification accuracy by involving clothing and hair information and it is also robust to occlusions and noise.

6. Conclusions and future work

In this paper, we have proposed a gender classification framework which utilizes not only facial information, but also hair and clothing. We extract features on facial components instead of on the whole face, which gives robustness against occlusions, illumination changes and noise. We prove that clothing information has discriminative ability, and design feature representations for hair and clothing information that is discarded in most existing work due to high variability. Moreover, classifier combination mechanisms are used to integrate various features to successfully boost the gender classification performance. Various experiments on FERET, BCMI and AR data sets are conducted to evaluate the performance of our framework. The experimental results show that our proposed framework improves classification accuracy, even in presence of occlusions and noise.

Our framework has two problems. First, hair feature extraction will be affected by a complex background. This can be solved by applying state-of-the-art segmentation algorithms. Second, computation time for hair feature extraction is a big bottleneck which impedes real-time application of our system. The future work is to accelerate this process.

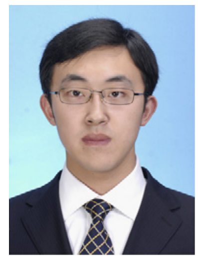
Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 60773090 and 90820018), the National Basic Research Program of China (Grant No. 2009 CB320901), the National High-Tech Research Program of China (Grant No. 2008AA02Z315), the Science and Technology Commission of Shanghai Municipality (Grant No. 09511502400), and the European Union Seventh Framework Programme (Grant No. 247619).

References

[1] B. Golomb, D. Lawrence, T. Sejnowski, Sexnet: a neural network identifies sex from human faces, Proceedings of Advances in Neural Information Processing Systems 3 (1990) 572–579.
[2] R. Brunelli, T. Poggio, HyperBF networks for gender classification, in: Proceedings of the DARPA Image Understanding Workshop, 1992, pp. 311–314.
[3] S. Gutta, H. Wechsler, Gender and ethnic classification of human faces using hybrid classifiers, in: Proceedings of International Joint Conference on Neural Networks, 1999, pp. 4084–4089.
[4] B. Moghaddam, M. Yang, Learning gender with support faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 707–711.
[5] H. Lian, B. Lu, Multi-view gender classification using multi-resolution local binary patterns and support vector machines, International Journal of Neural Systems 17 (2007) 479–487.
[6] H. Kim, D. Kim, Z. Ghahramani, S. Bang, Appearance-based gender classification with gaussian processes, Pattern Recognition Letters 27 (2006) 618–626.
[7] S. Baluja, H. Rowley, Boosting sex identification performance, International Journal of Computer Vision 71 (2007) 111–119.

[8] B. Xia, H. Sun, B. Lu, Multi-view gender classification based on local Gabor binary mapping pattern and support vector machines, in: Proceedings of International Joint Conference on Neural Networks, 2008, pp. 3388–3395.
[9] M. Erno, R. Roope, An experimental comparison of gender classification methods, Pattern Recognition Letters 29 (2008) 1544–1556.
[10] X. Lian, B. Lu, Gender classification by combining facial and hair information, in: Proceedings of Advances in Neuro-Information Processing, 2009, pp. 647–654.
[11] S. Josef, Z. Andrew, Video google: a text retrieval approach to object matching in videos, in: Proceedings of International Conference on Computer Vision 2, 2003, pp. 1470–1478.
[12] E. Brown, D. Perrett, What gives a face its gender? Perception 22 (1993) 829–840.
[13] T. Wu, C. Lin, R. Weng, Probability estimates for multi-class classification by pairwise coupling, Journal of Machine Learning Research 5 (2004) 975–1005.
[14] M. Sugeno, Theory of Fuzzy Integrals and its Applications, Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.
[15] P. Phillips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face-recognition algorithms, Image and Vision Computing 16 (1998) 295–306.
[16] A. Martinez, R. Benavente, The AR Face Database, Technical Report 24, Technical report CVC, 1998.
[17] K. Ueki, H. Komatsu, S. Imaizumi, K. Kaneko, N. Sekine, J. Katto, T. Kobayashi, A method of gender classification by integrating facial, hairstyle, and clothing Images, in: Proceedings of the Seventeenth International Conference on Pattern Recognition, vol. 4, 2004, pp. 446–449.
[18] G. Cottrell, J. Metcalfe, EMPATH: face, emotion, and gender recognition using holons, Proceedings of Advances in Neural Information Processing Systems 3 (1990) 564–571.
[19] H. Abdi, D. Valentin, B. Edelman, A. O'Toole, More about the difference between men and women: evidence from linear neural network and the principal-component approach, Neural Computing 7 (1995) 1160–1164.
[20] S. Tamura, H. Kawai, H. Mitsumoto, Male/female identification from 8×6 very low resolution face images by neural network, Pattern Recognition 29 (1996) 331–335.
[21] Z. Sun, G. Bebis, X. Yuan, S. Louis, Genetic feature subset selection for gender classification: a comparison study, in: Proceedings of IEEE Workshop on Applications of Computer Vision, 2002, pp. 165–170.
[22] A. Jain, J. Huang, Integrating independent components and linear discriminant analysis for gender classification, in: Proceedings of International Conference on Automatic Face and Gesture Recognition, 2004, pp. 159–163.
[23] H. Lian, B. Lu, E. Takikawa, S. Hosoi, Gender recognition using a min–max modular support vector machine, Advances in Natural Computation (2005) 433–436.
[24] Y. Saatci, C. Town, Cascaded classification of gender and facial expression using active appearance models, in: Proceedings of Seventh International Conference on Automatic Face and Gesture Recognition, 2006, pp. 393–398.
[25] B. Wu, H. Ai, C. Huang, Lut-based adaboost for gender classification, in: Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication, 2003, pp. 104–110.
[26] N. Sun, W. Zheng, C. Sun, C. Zou, L. Zhao, Gender classification based on boosting local binary pattern, in: Proceedings of Third International Symposium on Neural Networks, vol. 2, 2006, pp. 194–201.
[27] A. Burton, V. Bruce, N. Dench, What's the difference between men and women? Evidence from facial measurement, Perception 22 (1993) 153–176.
[28] Z. Ji, X. Lian, B. Lu, Gender classification by information fusion of hair and face, State of the Art in Face Recognition (2009) 215–230.
[29] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, in: Proceedings of European Conference on Computer Vision, 2004, pp. 469–481.
[30] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971–987.
[31] A. Lapedriza, D. Masip, J. Vitria, Are external face features useful for automatic face classification?, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 3, 2005, pp. 151–157.
[32] Y. Su, S. Shan, X. Chen, W. Gao, C. Beijing, Hierarchical ensemble of global and local classifiers for face recognition, in: Proceedings of International Conference on Computer Vision, 2007, pp. 1–8.
[33] P. Viola, M. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2004) 137–154.
[34] T. Cootes, C. Taylor, D. Cooper, J. Graham, Active shape models-their training and application, Computer Vision and Image Understanding 61 (1995) 38–59.
[35] T. Murofushi, M. Sugeno, An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure, Fuzzy Sets and Systems 29 (1989) 201–227.
[36] C. Chang, C. Lin, LIBSVM: A Library for Support Vector Machines, Software. Available at ⟨http://www.csie.ntu.edu.tw/~cjlin/libsvm⟩, 2001.



Bing Li received the BS degree in computer science and engineering from Shanghai Jiao Tong University in 2009. He is currently a MS candidate in the Brain-like Computing and Machine Intelligence Group, Shanghai Jiao Tong University. His research interests include face recognition, gender classification and image retrieval. He has been a recipient of several awards, including best bachelor student thesis of computer science department in Shanghai Jiao Tong University 2009.

Xiaochen Lian received the BS degree in computer science and engineering from Shanghai Jiao Tong University in 2008. He is currently a MS candidate in the Brain-like Computing and Machine Intelligence Group, Shanghai Jiao Tong University. His research interests include image retrieval, supervised codebook learning and gender classification. He has been a recipient of several awards and fellowships, including Microsoft Young Fellow sponsored by Microsoft Research Asia and Morgan Stanley Scholarship.

Bao-Liang Lu is a professor of Computer Science and Engineering at Shanghai Jiao Tong University (SJTU). He received his B.S. degree in instrument and control engineering from Qingdao University of Science and Technology, China, in 1982, the M.S. degree in computer science and engineering from Northwestern Polytechnical University, China, in 1989, and the Dr. Eng. degree in electrical engineering from Kyoto University, Japan, in 1994. From 1982 to 1986, he was with the Qingdao University of Science and Technology. From April 1994 to March 1999, he was a Frontier Researcher at the Bio-Mimetic Control Research Center, the Institute of Physical and Chemical Research (RIKEN), Japan. From April 1999 to August 2002, he was a Research Scientist at the RIKEN Brain Science Institute. Since August 2002, he has been a full Professor at the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. His research interests include brain-like computing, neural network, machine learning, pattern recognition, brain computer interface, computational linguistics, and computational biology and bioinformatics. He is a senior member of the IEEE.