We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Gender Classification by Information Fusion of Hair and Face

Zheng Ji, Xiao-Chen Lian and Bao-Liang Lu
*Department of Computer Science and Engineering, Shanghai
Jiao Tong University 800 Dong Chuan Road,
Shanghai 200240, China*

1. Introduction

Various gender classification methods have been reported in the literature. These existing methods fall into two categories. The first kind of method is the appearance-based approach. Golomb *et al.* [1] used a two-layer neural network with 30 × 30 inputs and directly fed the scaled image pixels to the network without dimensionality reduction. Their database contains only 90 images with half male and half female facial images. Gutta *et al.* [2] used the mixture of experts combining the ensembles of radial basis functions (RBF) networks and a decision tree. Xu *et al.* [3] applied Adaboost to gender classification problem with the feature pools composed of a set of linear projections utilizing statistical moments up to second order. Wu *et al.* [4] also adopted Adaboost. Instead of using threshold weak classifiers, they used looking-up table weak classifiers, which are more general and better than simple threshold ones due to stronger ability to model complex distribution of training samples. Moghaddam and Yang [5] demonstrated that support vector machines (SVMs) work better than other classifiers such as ensemble of radial basis function (RBF) networks, classical RBF networks, Fisher linear discriminant, and nearest neighbor. In their experiments, the Gaussian kernel works better than linear and polynomial kernels. However, they did not discuss how to set the hyper-parameters for Gaussian kernel, which affect the classification performance. Kim *et al.* [6] applied Gaussian process technique to gender classification. The advantage of this approach is that it can automatically determine the hyper-parameters.Wu *et al.* [7] presented a statistical framework based on the 2.5D facial needle-maps which is a shape representation acquired from 2D intensity images using shape from shading (SFS). Saatci and Town [8] used an active appearance model (AAM) of faces to extract facial features and developed a cascaded structure of SVMs for gender classification. Lian and Lu applied min-max modular support vector machine to gender classification and developed a method for incorporating age information into task decomposition [9]. They also proposed a multi-resolution local binary pattern for dealing with multi-view gender classification probelms [10].

The second kind of approach is the geometrical feature based approach. The idea is to extract from faces geometric features such as distances, face width, and face length. Burton *et al.* [11] extracted point-to-point distances from 73 points on face images and used discriminant analysis as a classifier. Brunelli and Poggio [12] computed 16 geometric

features, such as pupil to eye brow separation and eye brow thickness, from the frontal images of a face and used HyperBF network as a classifier.

Fig. 1. Illustration of the important role of hair information for gender classification. The upper row denotes three facial images of female, whose hair regions have been discarded, and the lower row denotes the same three images with hair.

Most of the existing methods mentioned above, however, use only facial information. As we know, external information such as hair and clothing, also provide the discriminant evidence. Fig. 1 illustrates the benefit of incorporating hair information in gender classification task. The upper row shows the facial images whose hair regions were discarded. The lower row shows the corresponding original pictures. It is difficult for us to judge the gender when we only see the images in the upper row since their neutral-like faces. But, when both the facial information and the external information of hair as shown in the lower row are presented, we can easily make the decision.

Although external features are useful, their detection, representation, analysis, and application have seldom been studied in the computer vision community. Considering the important role of hair features in gender classification, we study hair feature extraction and the combination of hair classifier and face classifier in this Chapter. Given a facial image containing both hair and face, we first locate hair region and face region. We construct a geometric hair model (GHM) to extract hair features and use local binary pattern (LBP) to extract facial features. After performing these feature extraction, we train two different classifiers for each kind of features and then apply a classifier fusion model. The key issue of classifier fusion is to determine how classifiers interact with each other. In this study, we adopt fuzzy integral [13], which has the advantage of its automatical adaptation of degree of classifier interaction.

We conduct experiments on three popular facial image databases: AR, FERET and CAS-PEAL. The experimental results demonstrate that the combination of hair and face classifiers achieves much higher classification rate than hair classifier or face classifier along.

The rest of this Chapter is organized as follows: Section 2 describes the feature extraction process of hair and face. Section 3 introduces the fuzzy integral method. Section 4 gives experimental details. Section 5 is the conclusions.

2. Feature extraction

2.1 Hair feature extraction

Hair is a highly variable feature of human appearance. It perhaps is the most variant aspect of human appearance. Until recently, hair features have often been discarded in most of the gender classification systems. To our best knowledge, there are two different algorithms in the literature about hair feature representation. Yacoob *et al.* [14] developed a computational model for measuring hair appearance. They extracted several attributes of hair including color, split location, volume, length, area, symmetry, inner and outer hairlines, and texture. These attributes are organized as a hair feature vector. Lapedriza *et al.* [15] learned a model set composed by a representative set of image fragments corresponding to hair zones called building blocks set. The building blocks set is used to represent the unseen image as it is a set of puzzle pieces and the unseen image is reconstructed by covering it with the most similar fragments. By using this approach, the hair information is encoded and used for classification. We adopt the former method and modify it in this study. The overall process of hair feature extraction is illustrated in Fig. 2.

Fig. 2. The overall process of hair feature extraction.

Geometric Hair Model The basic symbols used in the geometric hair model are depicted in Fig. 3. Here G is the middle point between the left eye point L and the right eye point R, I is the point on the inner contour, O is the point on the outer contour, and P is the lowest point of hair region.

Hair Detection In this work, we assume that the facial images used are in frontal view. The hair detection process is illustrated in Fig. 4. The detection algorithm consists of the following four steps:

1. Locate three landmarks on each facial image shown in the second picture of Fig. 4. Two are centers of eyes and the other one is middle of hair. These three points facilitate hair region extraction. Currently we label them manually. These points can be easily located in an automated manner, providing that the locations of the eyes are given.
2. Obtain binary facial image. The pixels around the landmark in hair region form the seeds to separate hair from face and background.
3. Get the edge image of the hair by Laplace operator to hair edge detection. The 2D Laplace operator is $\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$, and the edge extraction result is shown in the lower right image of Fig. 4.
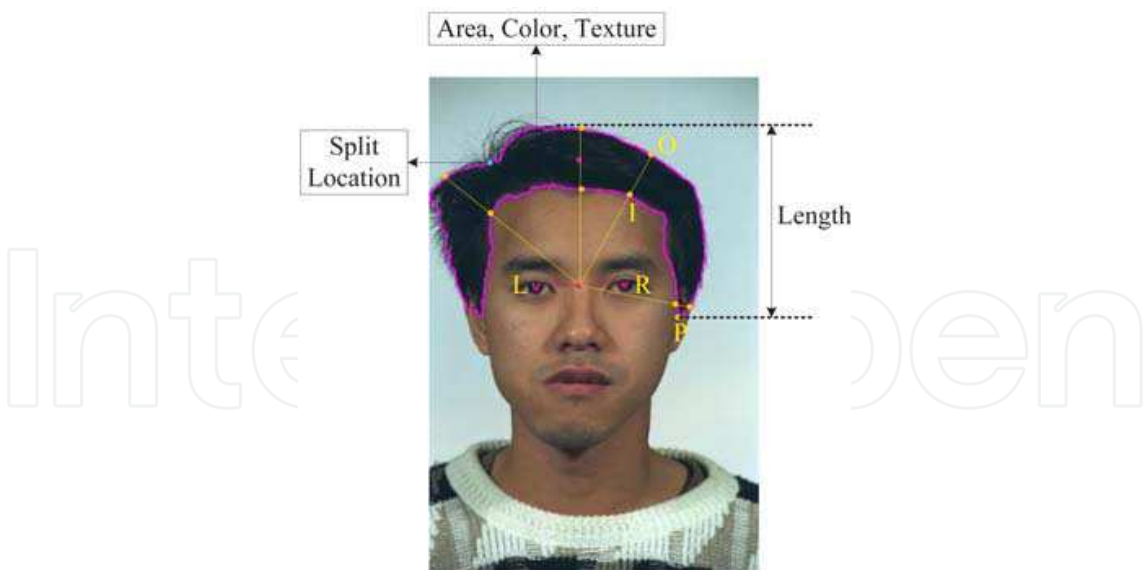
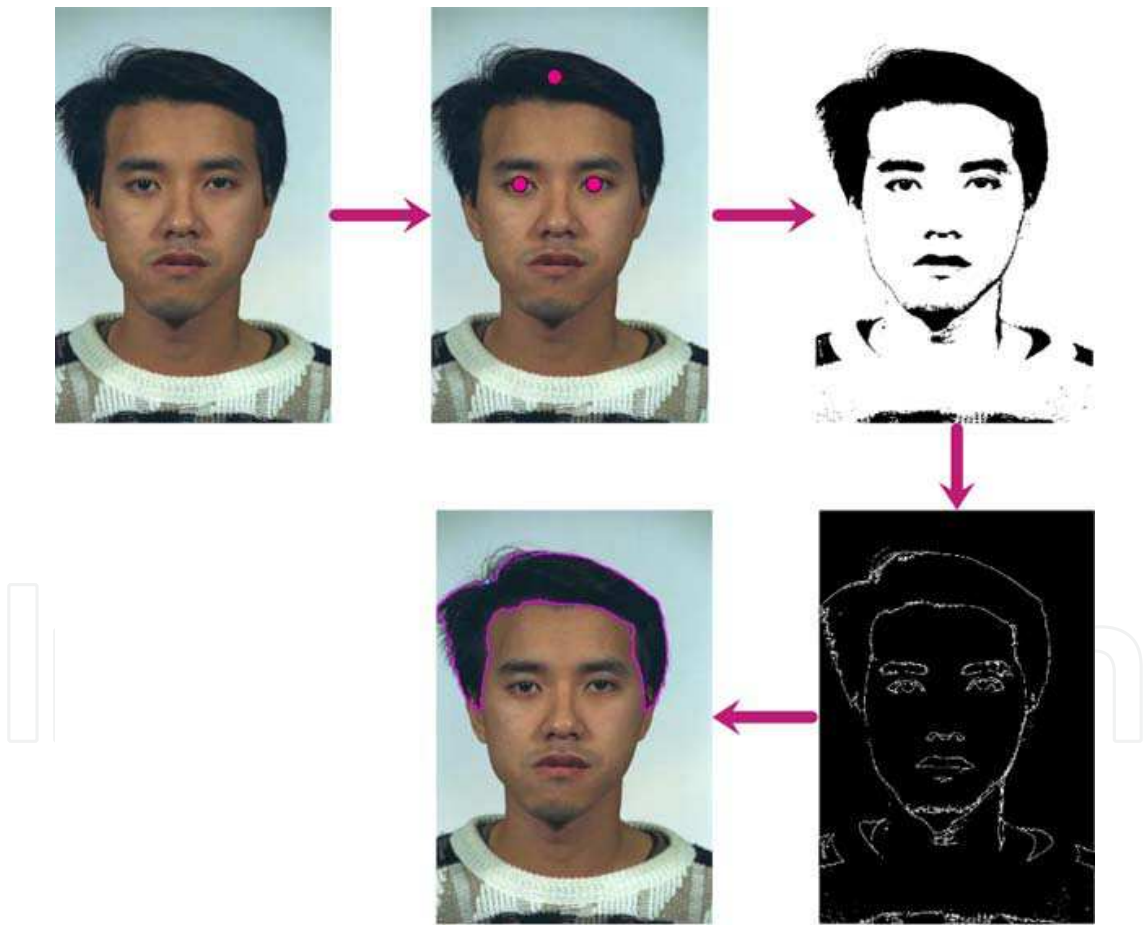Fig. 3. Some key parameters in geometric hair model.

Fig. 4. The process of hair detection.

4. Extract the inner and outer contour of the hair. Given an edge image from step 3), a ray denoted by the lines (yellow) in Fig. 3 is emitted from the mid-point of left and right eyes. In the ideal situation, the ray will meet exactly two edge points, I and O depicted in Fig. 3, respectively. By making a full rotation of the ray, we can determine the

contours of hair. In practice, however, step 2) may produce some holes in hair region, which make the detected contour inaccurate. To overcome this problem, we notice that the holes that will greatly affect the contour accuracy are those far away from the real contour and the distance between points *I* and *O* of consecutive rays will not change sharply. Based on this observation, we select the last edge point that the ray meets as the outer contour point. By using this technique, the holes can be removed. As a result, the detected contour of hair becomes more accurate. This improvement is illustrated in Fig. 5.
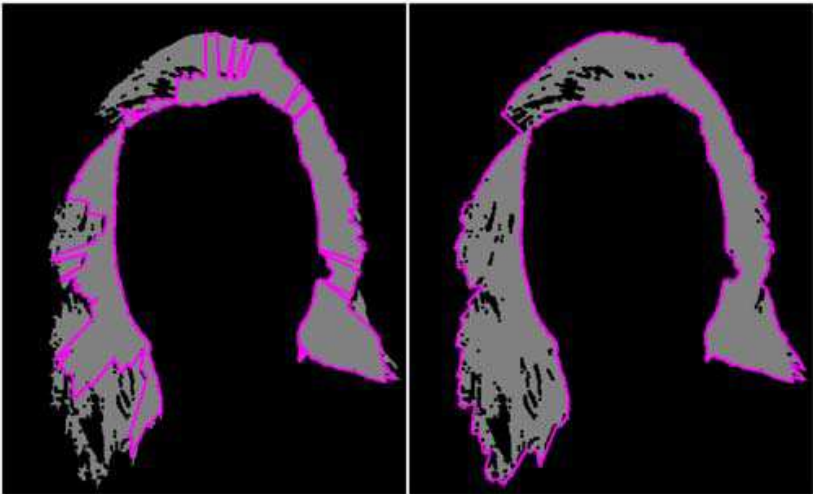
Fig. 5. The original detected contour of hair (left) versus the detected contour in which the holes have been removed (right).

Hair Length and Area We define the largest distance between a point on the outer contour and *P* as the hair length. The normalized distance L_{hair} is defined as

$$L_{hair} = max(dist(O_y, P_y))/Girth_{face} \tag{1}$$

where $Girth_{face}$ is the girth of the face region.

We define the area covered by hair as the hair surface. Based on the hair model, the normalized hair area is defined as

$$Area_{hair} = R\_Area_{hair}/R\_Area_{face} \tag{2}$$

where R_Area_{hair} is the real area of hair and R_Area_{face} is the area of face.

Hair Color To obtain the color in the hair region, we follow the method described in [16]. Based on this color model depicted in Fig. 6, the measured color results from the brightness and surface spectral reflectance. The averaged color distortion is calculated by

$$\overline{CD} = \frac{\sum_{i \in H} \|I_i - \alpha_i E_i\|}{|H|} \tag{3}$$

where *H* is the pixel set of hair region, and I_i and E_i denote the actual RGB color and the expected RGB color at pixel *i*, respectively, as follows:

$$I_i = (I_r(i), I_g(i), I_b(i)) \tag{4}$$
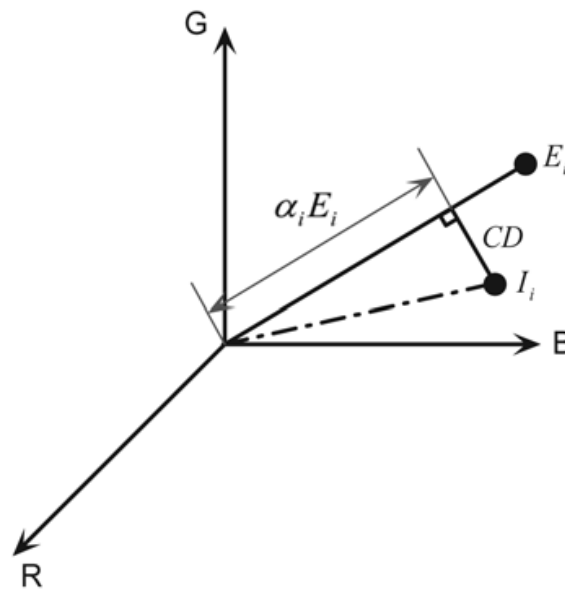
$$E_i = (E_r(i), E_g(i), E_b(i)) \tag{5}$$

Fig. 6. Color model of hair.

According to the definitions mentioned above, the color distortion CD_i at pixel i can be computed by

$$CD_i = \sqrt{(\frac{I_r(i) - \alpha_i \mu_r}{\sigma_r})^2 + (\frac{I_g(i) - \alpha_i \mu_g}{\sigma_g})^2 + (\frac{I_b(i) - \alpha_i \mu_b}{\sigma_b})^2} \tag{6}$$

where α_i represents the current brightness with respect to the brightness of the model

$$\alpha_i = \frac{(I_r(i)\frac{\mu_r}{\sigma_r})^2 + (I_g(i)\frac{\mu_g}{\sigma_g})^2 + (I_b(i)\frac{\mu_b}{\sigma_b})^2}{(\frac{\mu_r}{\sigma_r})^2 + (\frac{\mu_g}{\sigma_g})^2 + (\frac{\mu_b}{\sigma_b})^2}, \tag{7}$$

and (μ_r, μ_g, μ_b) and (σ_r, σ_g, σ_b) are the mean and standard deviation of color in the training set, respectively.

Hair Texture We employ Gabor wavelets to compute the hair feature attributes that characterize hair texture. The following two-dimensional Gabor function $g(x, y)$ under 6 directions and 4 scales and its Fourier transform $G(u, v)$ are used,

$$g(x,y) = (\frac{1}{2\pi\sigma_x\sigma_y})exp[-\frac{1}{2}(\frac{x^2}{\sigma_x} + \frac{y^2}{\sigma_y}) + 2\pi jW_x] \tag{8}$$

and

$$G(u,v) = exp-\frac{1}{2}[(\frac{u - W}{\sigma_u})^2 + (\frac{v}{\sigma_u})^2] \tag{9}$$

where $\sigma_u = \frac{1}{2\pi\sigma_x}$ and $\sigma_v = \frac{1}{2\pi\sigma_y}$.

The discrete form of Gabor wavelet transformation is defined as

$$W_{mn} = \iint I(x_l, y_l)g^*_{mn}(x - x_l, y - y_l)dx_l dy_l \tag{10}$$

where g^*_{mn} represents the conjugate operation in complex area of the *m*-th orientation and *n*-th scale.

We use the mean value and standard deviation of Gabor parameters to represent the texture shown in Fig. 7. The mean value and standard deviation are, respectively, calculated by

$$\mu_{mn} = \iint \mid W_{mn}(xy) \mid \tag{11}$$

and

$$\sigma_{mn} = \sqrt{\iint (\mid W_{mn}(x, y) - \mu_{mn} \mid)^2 dx dy}. \tag{12}$$
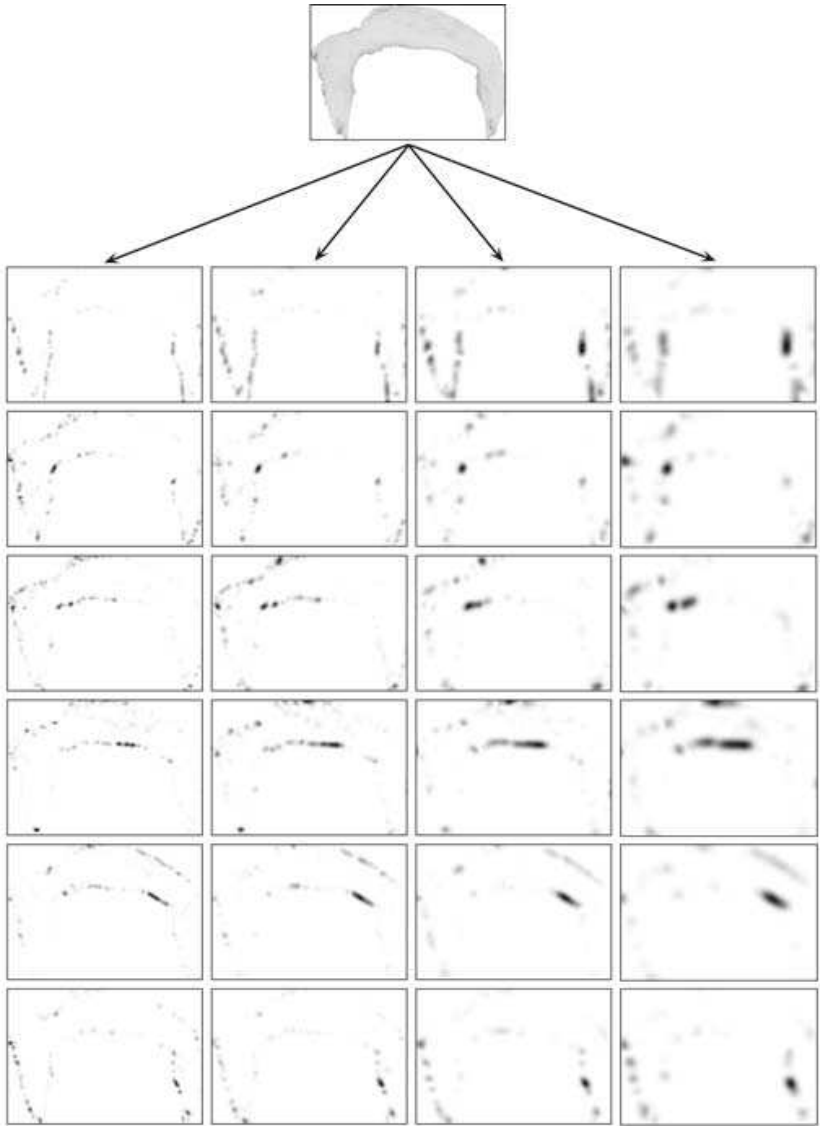

Fig. 7. Extracting hair texture using wavelet transformation with 4 scales and 6 orientations.

From Eqs. 11 and 12, we have the following feature attributes of hair texture:

$$V = [\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, \cdots, \mu_{24}, \sigma_{24}] \tag{13}$$

Hair-Split Location The hair split location is commonly accompanied by a concavity point at the outer hairline. The split angle is defined as the angle of the concavity point with respect to the horizontal axis. Whether a point P is a concavity point can be judged by

$$P = \begin{cases} concavity, & \overrightarrow{P_{n-1}P_n} \times \overrightarrow{P_nP_{n+1}} < 0 \\ non-concavity, & \overrightarrow{P_{n-1}P_n} \times \overrightarrow{P_nP_{n+1}} > 0 \end{cases} \tag{14}$$

We define *concavity*[P] as the concavity of point P, which can be calculated by

$$concavity[P] = -\frac{\overrightarrow{P_{n-1}P_n} \times \overrightarrow{P_nP_{n+1}}}{| \overrightarrow{P_{n-1}P_n} || \overrightarrow{P_nP_{n+1}} |} \tag{15}$$

With this definition, we scan all the edge points on the outer contour, compute the average concavity and select the one with largest *concavity*[P] as the split point.

By concatenating all the hair feature attributes mentioned above, we obtain a feature vector of hair as follows:

$$\begin{aligned} Hair_{Vector} &= [length, area, color, split\_location, texture] \\ &= [L_{hair}, Area_{hair}, E_r, E_g, E_b, concavity[P], \\ &\quad \mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, \cdots, \mu_{24}, \sigma_{24}] \end{aligned} \tag{16}$$

2.2 Face feature extraction

We use LBP [17] to characterize the face feature. The overall process is illustrated in Fig. 8. LBP is a simple and efficient approach for texture description. The operator labels the pixels of an image by thresholding 3 × 3-neighbourhoods of each pixel with the center value and considering the result as a binary number. The histogram of the labels is used as a texture descriptor. The basic LBP operator is illustrated in Fig. 9.

To achieve rotation invariance, An extension to the original operator is to use so called uniform pattern. A local binary pattern is called a uniform pattern 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, 00011110, and 10000011 are three uniform patterns.

We use the notation of $LBP_{P,R}^{u2}$ for the LBP operator. Here, $LBP_{P,R}^{u2}$ means using the operator in a neighborhood of P sampling points on a circle of radius R, and the superscript $u2$ represents using uniform patterns and labeling all remaining patterns with a single label. In our experiment, $\mathbf{LBP}_{8,1}^{u2}$ operator is used to quantify the total of 256 LBP values into the histogram of 59 bins according to the uniform strategy.

This histogram contains information about the distribution of the local micro-patterns over the whole image such as edges, spots and flat areas. For efficient face representation, one should also retain spatial information. For this purpose, an image is spatially divided into m small regions, $\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_m$, and the spatially enhanced histogram for region $\mathbf{R}_j$ is defined as

$$\mathbf{h}_{i,j} = \sum \mathbf{I}\{\mathbf{f}_l(\mathbf{x}, \mathbf{y}) = i\}\mathbf{I}\{(\mathbf{x}, \mathbf{y}) \in \mathbf{R}_j\}, \ i = 1, 2, ..., L, \ j = 1, 2, ..., m \qquad (17)$$

where L is the number of different labels produced by LBP operator, m is the number of blocks of the divided image, and $I\{A\}$ is 1 or 0 depending on whether A is true or false. According to Eq. 17, we obtain the following face feature vector:

$$\mathbf{H} = (\mathbf{h}_{1,1}, ..., \mathbf{h}_{L,1}, ..., \mathbf{h}_{1,m}, ..., \mathbf{h}_{L,m}) \qquad (18)$$
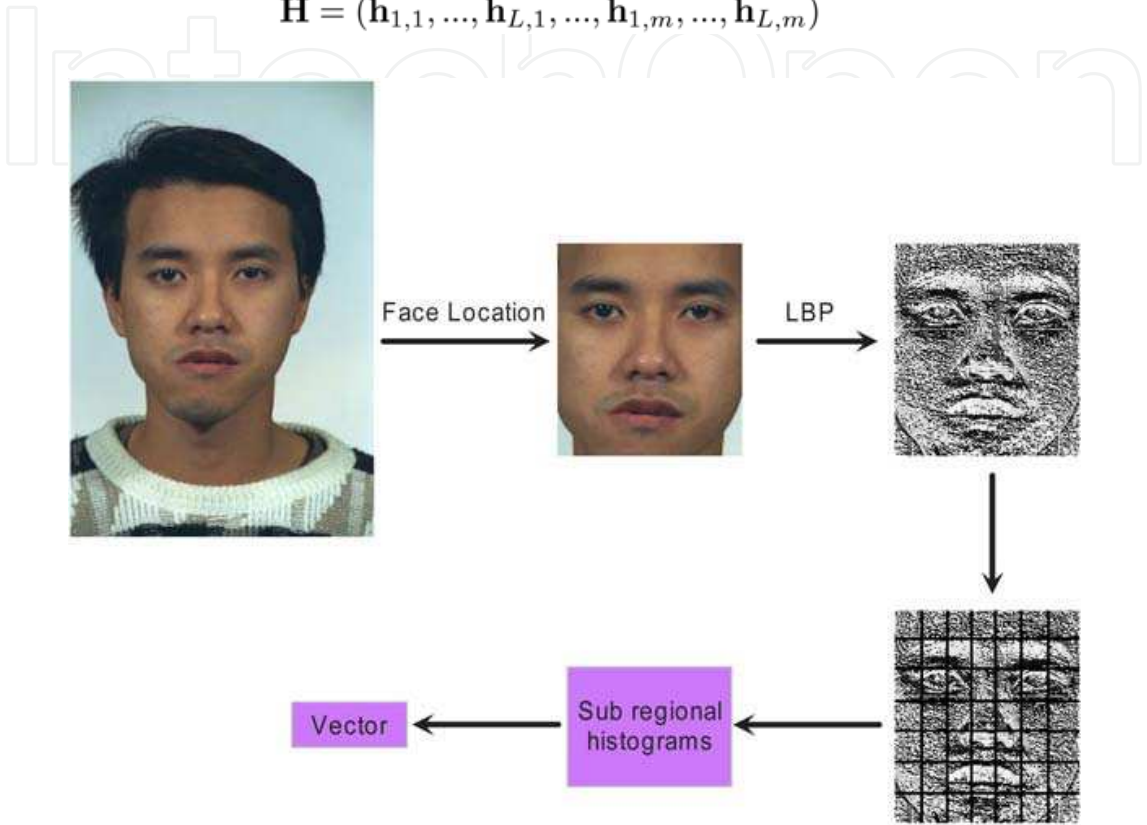

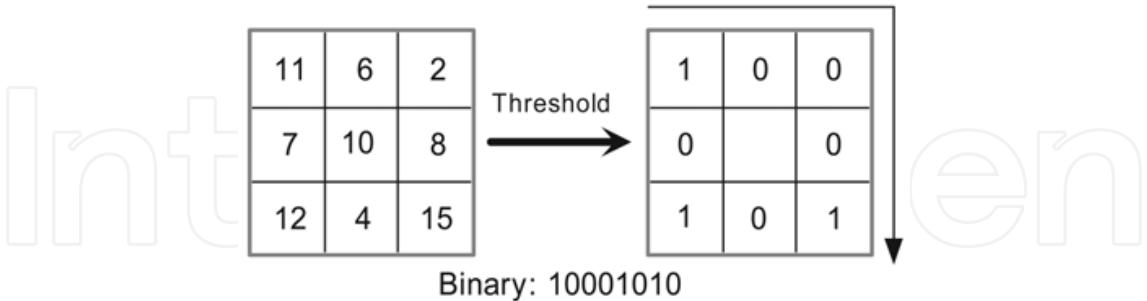Fig. 8. The overall process of LBP feature extraction.

Fig. 9. The 3 × 3-neighborhood LBP operator.

3. Fuzzy integral fusion of support vector machine classifiers

The ultimate goal of gender classification systems is to achieve the best possible classification performance. This objective traditionally led to the development of different classification schemes. Although the common way then is to choose one of the design that would yield the best performance, the sets of patterns misclassified by the different classifiers would not necessarily overlap. This suggested that different classifier designs

potentially offered complementary information about the patterns to be classified which could be harnessed to improve the performance of the selected classifier [18].

Here, we introduce a classifier fusion method based on fuzzy integral proposed by Sugeno [13]. The distinguish characteristic of fuzzy integral is that it is able to represent a certain kind of interaction between criteria, which is always avoided by making classifiers independent. Given a set of classifiers and their importance, fuzzy integral evaluates the interaction of these classifiers by computing fuzzy measure, a real function defined on the power set of classifiers. Based on such function and considering each classifier's decision, fuzzy integral will give a final decision.

Before describing the proposed fusion method, we present a note on notation. Let $C = \{c_1, \ldots, c_M\}$ be the set of classes and we have K classifiers, $f_1, \ldots, f_K$. Each of the K classifiers provides for an unknown sample X a degree of confidence in the statement 'X belongs to class c_j', for all c_j. We denote by $f_i^j(X)$ the confidence degree delivered by the classifier i of X belonging to c_j.

3.1 Probabilistic output of SVMs

We choose support vector machine as the basic classifier. Two SVMs are trained on hair and face features, respectively. Since fuzzy integral requires each classifier to give confidence value, we need to convert the binary output to probabilistic output. As the task of gender classification is a two-class problem, we assume that the training set is

$$\mathcal{T} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\} \tag{19}$$

where $x_i \in R_n$, $y_i \in \{-1, 1\}$, and $i = 1, 2, \cdots, N$.

A classifier $f(x)$ in the margin form of SVMs is equivalent to solving the following convex quadratic programming problem [19–21]

$$\begin{aligned} \min_{\mathbf{w}} \quad & \tfrac{1}{2}\|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{s.t. } & \forall 1 \leqslant i \leqslant N, y_i(\mathbf{w}^T\mathbf{x}_i + b) \geqslant 1 - \xi_i, \\ & \xi_i \geqslant 0 \end{aligned} \tag{20}$$

and its dual form:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i y_i \alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{N}\alpha_i \\ \text{s.t. } & \forall 1 \leqslant i \leqslant N, \quad 0 \leqslant \alpha_i < C, \\ & \sum_{i=1}^{N}\alpha_i y_i = 0 \end{aligned} \tag{21}$$

SVM classifier $f(x)$ finds the optimal hyperplane that correctly separates the training data while maximizing the margin. Therefore, there is the following discriminant hyperplane:

$$f(x) = \sum_{i=1}^{N}\alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{22}$$

where $K(\cdot, \cdot)$ is a kernel function and b is a bias.

Now $f(x)$ can be projected onto [0, 1] by sigmoid function

$$P_{+1}(x) = P(y = 1|x) = \frac{1}{1 + e^{Af(x)+B}}$$
$$P_{-1}(x) = 1 - P_{+1}(x)$$
(23)

The parameters A and B [22] can be estimated by solving the following maximum likelihood problem [23]:

$$\min_{A,B} F(A, B) = -\sum_{j=1}^{N}(t_i log(P_{+1}(x_i)) + (1 - t_i)log(P_{-1}(x_i)))$$

$$\text{s.t. } \forall 1 \leqslant i \leqslant N, t_i = \begin{cases} \frac{N_+ + 1}{N_+ + 2}, & if \ y_i = 1 \\ \frac{1}{N_- + 2}, & if \ y_i = -1 \end{cases}$$
(24)

where N_+ is the number of training samples with $y = 1$ and N_- is the number of training samples with $y = -1$.

3.2 Fuzzy integral theory

Fuzzy measures are the generalization of classical measures. The notion of a measure in an Euclidean space is a natural generalization of such elementary notions as the length of a line segment, the area of a rectangle and the volume of a parallelepiped. Still a more general concept of a measure in an arbitrary abstract set can be defined [24–26] as

Definition 1: By a measurable space we mean a pair (X,Ω) consisting of a set X and a σ-algebra of subsets of X. A subset A of X is called measurable (or measurable with respect to Ω) if $A \in \Omega$.

Definition 2: A measure μ on a measurable space (X,Ω) is a real non-negative set function defined for all sets of Ω such that $\mu(\emptyset) = 0$, and if $\{A_i\}_{i=1}^{\infty}$ is a disjoint family of sets with $A \in \Omega, i \geq 1$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$
(25)

It can be shown that a measure μ has the following properties [24]:

1. $\mu(A) \leq \mu(B)$ if $A \subset B$.
2. If $\{A_i\}_{i=1}^{\infty}$ is an increasing sequence of measurable sets, then

$$\lim_{i \to \infty} \mu(A_i) = \mu(\lim_{i \to \infty} A_i).$$
(26)

An important example of such a measure is the probability measure P, where $P(X) = 1$.

In the seventies of the twentieth century, alternative models were proposed by different researchers [27–30], who all share the following intuitively reasonable axioms:

Definition 3*: Let $g : \Omega$ [0, 1] be a set function with
1. $g(\emptyset) = 0, g(X) = 1,$

2. $g(A) \leq g(B)$ if $A \subset B$,
3. If $\{A_i\}_{i=1}^{\infty}$ is an increasing sequence of measurable sets, then

$$\lim_{i \to \infty} g(A_i) = g(\lim_{i \to \infty} A_i). \tag{27}$$

Such a function is called a fuzzy measure [29].

By the nature of the definition of a fuzzy measure g, the measure of the union of two disjoint subsets cannot be directly computed from the component measures. In light of this, Sugeno [29] introduced the so-called g_λ-fuzzy measures satisfying the property as follows:

$$\forall A, B \subset X, \; A \cap B = \emptyset, \\ \text{s.t.} \; g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B), \; \lambda > -1. \tag{28}$$

A g_λ-fuzzy measure is indeed a fuzzy measure, and the g_λ-fuzzy measure for $\lambda = 0$ is a probability measure [26, 31].

The constant λ can be determined by solving the following equation:

$$\lambda + 1 = \prod_{i}^{n}(1 + \lambda g^i) \tag{29}$$

In which, for a fixed set of $\{g^i = g_\lambda(s_i), 0 < g^i < 1\}$ where s_i is the ith classifier and g^i is a fuzzy density, there exists a unique root of $\lambda > -1$, $\lambda \neq 0$.

Let $S = \{s_1, s_2, \cdots, s_K\}$ be a finite set of individual SVM classifiers and $0 \leq h(s_1) \leq h(s_2) \leq \cdots \leq h(s_K) \leq 1$, where $h(s_i)$ is the probabilistic output of SVM classifier s_i ($1 \leq i \leq K$) and the Choquet integral can be computed by

$$\int_A h(s)dg(\cdot) = \sum_{i=1}^{K}[h(s_i) - h(s_{i-1})]g_\lambda(A_i) \tag{30}$$

where $A_i = \{s_1, s_2, ..., s_i\}$, $i = 1, 2, ..., K$, ($h(s_0) = 0$). Therefore, the value of $g_\lambda(A_i)$ can be computed recursively by

$$g_\lambda(A_1) = g_\lambda(s_1) = g^1 \\ g_\lambda(A_i) = g^i + g_\lambda(A_{i-1}) + \lambda g^i g_\lambda(A_{i-1}) \tag{31}$$

3.3 Fusion of SVM classifiers

Now for each f_k, the degree of importance g^k, of how important f_k is in the recognition of the class c_i must be determined. Hence λ can be calculated using Eq. (29) and the g_λ-fuzzy measure can be determined by a confusion matrix (CM) in Eq. (32) of the SVM classifier s_k which can be computed by means of n-flod validation method,

$$CM^k = \begin{bmatrix} n_{11}^k & n_{12}^k & \cdots & n_{1M}^k \\ n_{21}^k & n_{22}^k & \cdots & n_{2M}^k \\ \vdots & \vdots & \ddots & \vdots \\ n_{M1}^k & n_{M2}^k & \cdots & n_{MM}^k \end{bmatrix} \tag{32}$$

where n_{ij}^k is the number of the samples which are the class c_i assigned to the class c_j by the classifier s_k. Therefore, the g_λ-fuzzy measure can be calculated by the following formulation:

$$g^k = \frac{\sum_{i=1}^m n_{ii}^k}{\sum_{i=1}^m \sum_{j=1}^m n_{ij}^k} \tag{33}$$

and

$$g_i^k = \frac{n_{ii}^k}{\sum_{j=1}^m n_{ji}^k}, i = 1, 2, \cdots, m, k = 1, 2, \cdots, K \tag{34}$$

where g^k is the fuzzy density for the SVM classifier s_k and g_i^k represent the fuzzy density for the class c_i and the SVM classifier s_k. According to the above statement, we conclude the following algorithm for fuzzy integral fusion of SVM (FIF-SVM) classifiers.

Algorithm: Fuzzy integral fusion of SVM classifiers.
step 1: calculate λ using Eq. (29) after determining g^k or g_i^k in Eq. (34);
step 2: **for** each class c_i **do**
        for each SVM classifier s_k **do**
            calculate $h_i(s_k)$ using Eqs. (22) and (23);
            determine $g_i(A_k)$ by Eq. (32);
        end_for
        compute the fuzzy integral $f(c_i)$ for the class c_i by Eq. (30);
    end_for
step 3: Output $argmax\{f(c_1), f(c_2), \cdots, f(c_M)\}$

4. Experiments

4.1 Databases
In this work, a total number of 2608 frontal facial images were selected from three popular face databases. Among them, 481 male and 161 female images were selected from the AR database randomly; 595 male and 445 female frontal facial images are chosen from the CAS-PEAL face database [32]; and 583 male and 406 female facial images are selected from the FERET face database. These facial images are described in Table 1.

Database	Total data	Male	Female	Training	Test
AR	579	481	161	128*2	323
CAS-PEAL	1040	595	445	356*2	328
FERET	989	583	406	324*2	341
Total	2608	1596	1012	808*2	992

Table 1. The data set.

4.2 Fuzzy measure values
In Table 2, the fuzzy measure values of all the classes based on face feature are larger than those of hair feature. However, in our experiments, the fuzzy measure depends on the confusion matrix which can be acquired by 5-fold validation after training. The classifier which takes on good classification performance after cross validation can achieve a high fuzzy measure for each class. The reason is that the confusion matrix is a detailed description form of classification accuracy. Because the classification precision of the gender

classifier based on face feature is higher than hair feature in the course of cross validation, the fuzzy measure favors the SVM (face) classifier. Generally speaking, there are different mapping parameters, *A* and *B* in Eq. (23), for different classifiers, which are independent in the process of training and test mutually. At the same time, there is $\lambda > -1$ in most situations. However, there exists $\lambda = -1$ in Table 2 because of some classification accuracy values equal to 100% after cross validation. Therefore, if the classification precision is less than 100%, there exists $\lambda > -1$.

No.	Database	Kernel	g^1 (hair)		g^2 (face)		A		λ	
			male	female	male	female	hair	face	male	female
1		Linear	0.8931	0.9120	1.0000	1.0000	-1.9693	-2.5027	-1.0000	-1.0000
2	AR	Poly	0.9370	0.9302	1.0000	0.9922	-3.1204	-4.9165	-1.0000	-0.9994
3		RBF	0.9297	0.9297	1.0000	1.0000	-2.8025	-5.5546	-1.0000	-1.0000
4		Linear	0.8391	0.8859	0.9581	0.9633	-0.4496	-1.7323	-0.9916	-0.9951
5	CAS-PEAL	Poly	0.8711	0.8732	0.9583	0.9688	-0.4274	-3.7717	-0.9936	-0.9953
6		RBF	0.8837	0.8587	0.9583	0.9688	-0.5357	-4.2729	-0.9943	-0.9947
7		Linear	0.8864	0.8701	0.9179	0.9310	-0.5231	-3.6407	-0.9885	-0.9889
8	FERET	Poly	0.9025	0.8879	0.9067	0.9574	-0.3453	-4.1190	-0.9889	-0.9944
9		RBF	0.9143	0.8919	0.9196	0.9519	-0.8303	-7.5197	-0.9918	-0.9939

Table 2. Values of g_λ-fuzzy measure, parameter A ($B = 0$) and λ.

4.3 Classification results

In our experiments, both the proposed geometric hair model and the LBP approach are used to extract hair and face features, respectively, from all the training samples. Two kinds of SVM classifiers, namely hair classifier and face classifier, are trained on the given data sets shown in Table 1. The results of *five*-fold validation are employed to calculate the confusion matrix, which determines the g_λ-fuzzy measure values of these two SVM classifiers. The classification accuracy of three different classifiers on the test data is described in Table 3.

No.	Database	Kernel	Hair	Face	Face & Hair
1		Linear	84.52	83.90	**93.19**
2	AR	Poly	83.59	89.16	**93.19**
3		RBF	84.52	85.45	**93.50**
4		Linear	89.02	96.34	**96.95**
5	CAS-PEAL	Poly	90.55	96.34	**97.87**
6		RBF	88.72	96.95	**98.48**
7		Linear	81.23	93.26	**94.43**
8	FERET	Poly	84.16	93.26	**94.13**
9		RBF	84.16	92.67	**94.13**

Table 3. Accuracy (%) of three different classifiers: face SVM-classifier along, hair SVM-classifier along, and fuzzy integral fusion of face and hair SVM-classifiers.

From Table 3, we can see that the proposed fuzzy integral fusion method achieves the best classification accuracy in all the cases. It should be noted that when the classification accuracy of both hair classifier and face classifier are relatively lower, the proposed fusion method can dramatically improve the classification accuracy. At the same time, we can see that the face classifier has higher classification accuracy than that of the hair classifier. This indicates that internal features such as face feature are more critical to gender classification than external features such as hair feature. On the other hand, hair features play a good complementary role for gender classification.

5. Conclusions

We have presented a modified geometric hair model for extracting hair features for gender classification. By using this model, hair features are represented as length, area, color, split-location, and texture. In order to integrate the outputs of both hair classifier and face classifier that use hair features and face features, respectively, we have proposed a classifier fusion approach based on fuzzy integral theory. The experimental results on three popular face databases demonstrate the effectiveness of the modified geometric hair model and the proposed classifier fusion method. From the experimental results, we can obtain the following observations. a) Hair features play an important role in gender classification; b) Face features are more critical than hair features to gender classification; c) Implementing the fusion of hair and face classifiers can achieve the best classification accuracy in all of the cases; d) The proposed fusion method can improve the classification accuracy dramatically when the performance of all the single classifier is not good. From this study, we believe that more external features such as hair and clothes should be integrated into face features to develop more reliable and robust gender classification systems.

6. Acknowledgments

7. References

[1] Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: Sexnet: A neural network identifies sex from human faces. In: Advances in neural information processing systems 3, Morgan Kaufmann Publishers Inc. 572–577 (1990)
[2] S. Gutta, J.R.J. Huang, P.J., Wechsler, H.: Mixture of experts for classification of gender, ethnic origin, and pose of human faces. IEEE Trans. Neural Networks 11(4) 948–960 (2000)
[3] Xu, X.H., S., T.: Soda-boosting and its application to gender recognition. Lecture Notes in Computer Science (4778) 193–204 (2007)
[4] Wu, B., Ai, H., Huang, C.: Lut-based adaboost for gender classification. In: AVBPA03. 104–110 (2003)
[5] Moghaddam, B., Yang, M.: Learning gender with support faces. IEEE Trans. on Pattern Analysis and Machine Intelligence 24(5) 707–711 (May 2002)
[6] Kim, H., Kim, D., Ghahramani, Z., Bang, S.: Appearance-based gender classification with gaussian processes. Pattern Recognition Letters 27(6) 618–626 (April 2006)
[7] Wu, J., Smith, W., Hancock, E.: Gender classification using shape from shading. In: BMVC07. xx–yy (2007)
[8] Saatci, Y., Town, C.: Cascaded classification of gender and facial expression using active appearance models. In: FGR06. 393–400 (2006)
[9] Lian, H.C., Lu, B.L.: Gender recognition using a min-max modular support vector machine. The First International Conference on Natural Computation 433–436 (2005)
[10] Lian, H.C., Lu, B.L.: Multi-view gender classification using multi-resolution local binary patterns and support vector machines. International Journal of Neural System 17(6) 479–487 (2007)

[11] Burton, A., Bruce, V., Dench, N.: What's the difference between men and women? Evidence from facial measurement. Perception 22(2) 153–76 (1993)

[12] R.Brunelli, Poggio, T.: Hyperbf networks for gender classification. In: DARPA Image Understanding Workshop. 311–314 (1992)

[13] Sugeno, M.: Theory of fuzzy integrals and its applications. PhD thesis, Tokyo Institute of Technology (1974)

[14] Yacoob, Y.: Detection and analysis of hair. IEEE Trans. Pattern Anal. Mach. Intell. 28(7) 1164–1169 (2006) Member-Larry S. Davis.

[15] Lapedriza, A., Masip, D., Vitria, J.: Are external face features useful for automatic face classification? In: CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) -Workshops, Washington, DC, USA, IEEE Computer Society 151 (2005)

[16] Yacoob, Y., Davis, L.: Detection, analysis and matching of hair. The tenth IEEE International Conference on Computer Vision 1 741–748 (2005)

[17] Ojala, T., Pietik¨aainen, M., M¨aaenp¨aa¨aaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. on Pattern Analysis and Machine Intelligence 24(7) 971–987 (2002)

[18] Kittler, J., Hatef, M., Duin, R., Matas, J.: On combining classifiers. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20(3) 226–239 (1998)

[19] Cristianini, N., Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods. Publishing House of Electronics Industry (2004)

[20] Sloin, A., Burshtein, D.: Support vector machine training for improved hidden markov modeling. IEEE Transactions on Signal Processing 56(1) 172--188 (2008)

[21] Williams, P., Li, S., Feng, J., Wu, S.: A geometrical method to improve performance of the support vector machine. IEEE Transactions on Neural Networks 18(3) 942--947 (2007)

[22] Platt, J.: Probabilistic outputs for support vector machines and comparison to regularized likelihood method. In Advance in Large Margin Classifiers, A. Smola, P. Bartlett, B. Scholkopf, D. Schuurmans, Eds, Cambridge, MA:MIT Press (2000)

[23] Zhang, Y.Z., Liu, C.M., Zhu, L.K., Hu, Q.L.: Constructing multiple support vector machines ensemble based on fuzzy integral and rough reducts. The 2nd IEEE Conference on Industrial Electronics and Applications 1256–1259 (2007)

[24] Halrnos, P.R.: Measure theory. New York:Van Nostrand (1950)

[25] Pfeffer, W.F.: Integrals und meusures. New York:Marcel Dekker (1977)

[26] Tahani, H., Keller, J.M.: Information fusion in computer vision using the fuzzy integral. IEEE Transactions on Systems, Man and Cybernetics 20(3) 733–741 (1990)

[27] Shafer, G.A.: A mathematical theory of evidence. Princeton, NJ:Princeton Univ. Press (1976)

[28] Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. int. J. Fiizzy Sets Syst. 1(1) 3–28 (1978)

[29] Sugeno, M.: Fuzzy measures and fuzzy integrals: A survey. Fuzzy Automata and Decision Processes, Amsterdam: North Holland 89–102 (1977)

[30] Terano, T., Sugeno, M.: Conditional fuzzy measures and their application. In Fuzzy Automata and Tlzrir Applications to Cognitive and Decision Processes, New York:Academic Press 151–170 (1975)

[31] Banon, G.: Distinction between several subsets of fuzzy measures. Fuzzy Sets Syst. 5 291–305 (1981)

[32] Gao, W., Cao, B., Shan, S.G., et al.: The cas-peal large-scale chinese face database and baseline evaluations. Technical report of JDL, available on http://www.jdl.ac.cn/~peal/peal tr.pdf (2004)
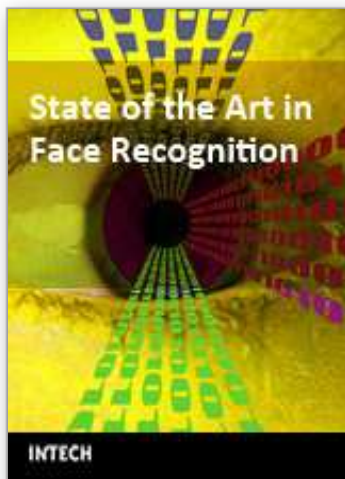
State of the Art in Face Recognition

Edited by Julio Ponce and Adem Karahoca

Notwithstanding the tremendous effort to solve the face recognition problem, it is not possible yet to design a face recognition system with a potential close to human performance. New computer vision and pattern recognition approaches need to be investigated. Even new knowledge and perspectives from different fields like, psychology and neuroscience must be incorporated into the current field of face recognition to design a robust face recognition system. Indeed, many more efforts are required to end up with a human like face recognition system. This book tries to make an effort to reduce the gap between the previous face recognition research state and the future state.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Zheng Ji, Xiao-Chen Lian and Bao-Liang Lu (2009). Gender Classification by Information Fusion of Hair and Face, State of the Art in Face Recognition, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-42-4, InTech, Available from:
http://www.intechopen.com/books/state_of_the_art_in_face_recognition/gender_classification_by_information_fusion_of_hair_and_face

INTECH
open science | open minds