Neurocomputing 175 (2016) 515–528

Contents lists available at ScienceDirect

Neurocomputing

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

Semantic description method for face features of larger Chinese ethnic groups based on improved WM method

Yuangang Wang [a,b], Xiaodong Duan [b,*], Xiaodong Liu [a], Cunrui Wang [b,c], Zedong Li [b,c]

[a] Research Center of Information and Control, Dalian University of Technology, Dalian, Liaoning, China
[b] Dalian Key Lab of Digital Technology for National Culture, Dalian Nationalities University, Dalian, Liaoning, China
[c] Institute of System Science, Northeastern University, Shenyang, Liaoning, China

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form
26 October 2015
Accepted 26 October 2015
Communicated by D. Wang
Available online 1 November 2015

Keywords:
Chinese ethnic groups
Improved WM Method
Anthropometry
Morphology
Facial geometric features

ABSTRACT

In this paper, a semantic description method based on improved WM algorithm is proposed to characterize the facial features of larger Chinese ethnic groups. We firstly utilize the face landmarking technique to extract facial feature points automatically. Geometric features are defined with these detected landmarks, including distances, perimeters and areas. Then the WM method is improved to generate linguistic rule from facial geometric feature data, which implements semantic description for multi-ethnic facial characteristics. Finally, a case study of learning ethnicity from face with proposed method is investigated in CEFD database. The experiment results indicate that the linguistic rule base obtained by method is competitive in ethnicity recognition compared with method Naive Bayes, C4.5, Decision Table, Random Forest, Adaboost and Logistic regression in terms of accuracy and interpretability.

1. Introduction

Due to differences in inheritance, geographical environment and cultural background, every ethnic group in China is unique in their facial features. However, the dissimilarity is shrinking result from intermarriage and amalgamation between nationalities. So the research on multi-ethnic morphology plays a significant role in exploring the interrelation of different ethnic groups and preserving facial morphological diversity. Moreover, craniofacial reconstruction may depend on analysis of facial features, especially the minority people. Anthropometric methods are introduced into clinical practice to quantify changes in the craniofacial framework and features distinguishing various ethnic groups are discovered in the field of anthropology [1–4]. The classic measuring methods are manually implemented in subjects' face, high-quality images or X-ray pictures, which causes the defects in efficiency, experiment scale and practical application. For analyzing multi-ethnic facial features in China, the traditional methods must be replaced by time-saving and computer-assisted ones.

As a model-free and data-driven tool in soft computing, FRBSs (Fuzzy Rule-based Systems) are initially proposed to model nonlinear and complex systems in the field of intelligent control. During the past few decades, FRBSs are extensively applied in several fields. In FRBSs, fuzzy IF–THEN rules play a crucial role, which directly affect the systems' performance in accuracy and interpretability. There are lots of approaches for extracting fuzzy rules from data, such as heuristic methods, and methods based on genetic algorithms, fuzzy clustering, and neural networks, among others. The WM (Wang and Mendel) method, proposed in [5,6], is a classic method capable of simplicity, efficiency and practicability. Since the WM method highly depends on samples, the extracted fuzzy rule base is poor in completeness and robustness if noise data exists in dataset. In addition, with the increase in scale of samples, the algorithm complexity will obviously improve while the efficiency will decline sharply. For these weaknesses, considerable improvements on WM methods are investigated in [7–10]. Furthermore, the WM method is utilized to tackle many practical problems, such as friction modelling and control compensation [11], electric loading prediction [9], transient identification in nuclear power plant [12] and big data analysis [13,14]. What's more, the WM method is also employed to facial expression recognition [15,16] and face detection [17]. But there are rare literatures about analyzing ethnic/racial face feature by WM method.

In this work, we focus on how to effectively perform semantic description for Chinese multi-ethnic face features. Geometrical features are extracted based on anthropology research results by face landmarking. WM method is improved to mine the IF–THEN rules from geometrical features for every ethnic group, which are used for learning ethnic group from face. The rest of this paper is organized as follows. In Section 2, we briefly present the applied tools and techniques including face landmark and improved WM

* Corresponding author.
E-mail address: dxd_dlnu@163.com (X. Duan).

http://dx.doi.org/10.1016/j.neucom.2015.10.089
0925-2312/

method. Then in Section 3, a semantic description method for face features of larger Chinese ethnic groups based on improved WM method is proposed. Section 4 is the case study on learning ethnic groups from face by rule base generated in the previous section and experiment results compared with several popular classification algorithms. Section 5 concludes the paper and summarizes the merit and insufficiency to be improved.

2. Applied tools and techniques

2.1. Face landmarking

Face landmarking, defined as the detection and localization of certain characteristic points on the face, is an important intermediary step for many subsequent face processing operations that range from biometric recognition to the understanding of mental states [18]. Accurate face landmarking is significant step that plays a foundational role in facial geometric features extraction. Currently, existing landmarking methods can be mainly divided into model-based methods and texture-based ones. For the former, facial image and landmarks are viewed as a whole shape. They obtain "mean shapes" from hand-labelled training face images and then try themselves best to make the proper shape fit for the unknown face. The representative methods include ASM, AAM and neural network-based algorithms. For the latter, they are characterized independent of model trained in advance when finding landmark. So, texture-based methods are also known as none model-based methods. These methods are mostly based on transformation or template, such as Gabor filters, HOG features, principal component analysis (PCA), independent component analysis (ICA), Gabor transform, Kernel PCA (KPCA), and local linear embedding (LLE). In this paper, we adopt active shape models with SIFT Descriptors and MARS [19] as landmarking method because of its efficiency and performance against available techniques for automatic face landmarking on frontal faces.

Active Shape models (ASM) are statistical models of the shape of objects which iteratively deform to fit to an example of the object in a new image, developed by Tim Cootes and Chris Taylor [20]. It is one of the most widely used algorithms in face landmarking. Later, various descendants of ASM are continually proposed. Among various descendants of ASM, Milborrow and Nicolls [21] make significant contribution to ASM methodology. In addition to improvements on original algorithm, what is more practical is that corresponding software named standard ASM (STASM) is freely opened source under a BSD style license. Comparative study on face landmarking algorithms is conducted under databases (CK+, BioID) in [18], and the results indicate that STASM scores the best performance. So, it is widely adopted by many researchers.

The improvements of STASM largely lie in four aspects. Firstly, two-dimensional profiles are used at each landmark instead of one-dimensional one. Because two-dimensional profile includes more information around the landmark, and the information may boost the fitting results if used rationally. Secondly, loosening up the shape model is done by varying the number of shape eigenvectors in the progress of adapting the shape model. Thirdly, the calculation of Mahalanobis distance is a very time consuming process, which can be tackled by trimming the covariance matrix. Finally, running two ASM searches in series is adopted to decrease the bad effect if the start shape is not well positioned. A recent addition to the STASM is the study in [19], where the authors use a simplified form of scale invariant feature transform (SIFT) descriptors for template matching, replace the one-dimensional profile used in the classical model, utilize multivariate adaptive regression splines to efficiently match these descriptors around the landmark and introduce several techniques for heavily decreasing their computation cost so as to make SIFT based ASM more practical. In this paper, we adopt the active shape models with SIFT descriptors and MARS (stasm v4.1.0, http://www.milbo.users.sonic.net/stasm/) to extract face feature points, which lays a solid foundation for building geometrical features.

2.2. Improved WM method

From the view of machine learning or pattern recognition, WM method is often denoted as a classification algorithm [22]. Compared with other classification methods, WM is capable of providing relatively ideal accuracy and generating an interpretable model with linguistic labels. In the case of pattern classification, the training dataset consisted of N input–output data pairs:

$$T = \{(x^{(p)}, y^{(p)})\}, \quad p = 1, 2, \ldots, N \tag{1}$$

where $x^{(p)} = (x_1^{(p)}, \ldots, x_n^{(p)}) \in R_n$ represents attributes, $y^{(p)} \in \{c_1, \ldots, c_K\}$ denotes class labels, and N is the number of samples. The relationships between input and output can be described with IF–THEN rules by FRBSs. They are expressed as follows:

$$R^{(p)} : \textbf{IF } x_1^{(p)} \text{ is } A_{l_1}^{x_1}, \ldots, x_n^{(p)} \text{ is } A_{l_n}^{x_n} \textbf{ THEN } y^p \text{ is } c^{(p)} \tag{2}$$

where $x_i^{(p)}$ is the ith attribute value of pth sample, $A_{l_j}^{x_i}$ is the jth fuzzy set of ith attribute, and $A(x_i) = \{A_{l_1}^{x_i}, \ldots, A_{l_j}^{x_i}\}$.

The main process of generating fuzzy rule base consists of the following steps:

Step 1: *The fuzzification of input variable.* In this work, input variable x_i is the attribute of sample. The domain of interval value for every variable is $[x_i^{min}, x_i^{max}]$, $i = 1, 2, \ldots, n$. In each domain, l_j fuzzy sets are defined, namely $A(x_i) = \{A_{l_1}^{x_i}, \ldots, A_{l_j}^{x_i}\}$. In general, l_j is assigned by 3 or 5 which indicates fuzzy sets $\{small, middle, large\}$ or $\{verysmall, small, middle, large, verylarge\}$. At last, every fuzzy division $A(x_i)$ is specified by a membership function. The alternative ones include triangular, trapezoidal and gaussian, etc.

Step 2: *Generating fuzzy rules from samples.* For every sample $x^{(p)}$, we can calculate the rule weight by

$$R_W(x^{(p)}) = \prod_{i=1}^{n} \sup_{l_j \in U} \mu_{A_{l_j}^{x_i}}(x_i^{(p)}) \tag{3}$$

where $U = \{small, middle, large\}$. At the same time, the attributes of sample $x^{(p)}$ are replaced by linguistic labels corresponding to implications of fuzzy sets. So, the rule base consisting of N rules can be generated from N samples.

Step 3: *Simplifying the rule base.* During the process of generating rule base, different samples may be same in their rule's antecedents. In this case, the rule with less weight will be removed from rule base.

However, the above-mentioned method is sensitive to noise data. For the practical situation in this paper, it is not persuasive that only rule weight is used for evaluating the rule's value. Inspired by [11], both the rule weight and the number of its covering samples should be considered to simplify the rule base. But the definition in [11] is designed for approximating the nonlinear function. In this paper, the rule's support degree for classification is defined as follows.

Definition 1. Let T be the dataset and M be a subset of T. For $m_p, m_q \in M$, $p, q \in \{1, 2, \ldots, |M|\}$, m_p and m_q is covered by the same rule r_M generated by M. The support degree of r_M can be calculated by

$$S(r_M) = \frac{\sum_{k \in M} \prod_{i=0}^{n} \sup_{j \in A(x_i)} \mu_{A_{l_j}^{x_i}}(x_i^{(k)})}{\sum_{k=1}^{|T|} \prod_{i=0}^{n} \sup_{j \in A(x_i)} \mu_{A_{l_j}^{x_i}}(x_i^{(k)})} \tag{4}$$

For the convenience of computing, we can utilize the following equation instead of (4).

$$S(r_M) = \frac{|M|}{|T|} \tag{5}$$

The implication of support degree calculated by (4) consisted of the number of covering samples and their rule weight, while (5) only considers the former. Eq. (5) is often used when the scale of dataset is large. The rule base can be simplified by support degree rather than rule weight, which may eliminate the effect caused by noise data or outlier as far as possible.

Step 4: *Classifying the sample.* In conventional WM method, for a sample to be classified, the association grade between sample and rules is calculated to determine which class the sample belongs to. The sample's class label is assigned by the consequent part of rule with highest association grade. Nevertheless, in practical application, the classification accuracy of this method is seriously affected by noise data and rule base's incompleteness. Concerning this issue, the nearest rule is defined as follows

Definition 2. Let R be a set of rules and X be a set of samples. For $x \in X$, $r_x \notin R$. For $r \in R$, M_r is the set of samples covered by rule r. C_r is called the center of rule r if $C_r = \arg\max_{x \in M_r} R_W(x)$, where R_W refers to (3). For $y \in X$, we can compute the Euclidean distance by

$$d(y, C_r) = \left(\sum_{j=1}^{n} (y^j - C_r^j)^2 \right)^{1/2} \tag{6}$$

r is called nearest rule if $C_r = \arg\min_{c \in C} d(y, c)$ is satisfied, where C is a set of rule center in R.

On the basis of above definition, the nearest rule can be found to classify the unclassified sample whose corresponding rule is not included in rule base. The nearest rule is viewed as an approximation for the rule generated by unclassified sample. The approximation is based on the intrinsic property of classification problem that a sample may be always classified into a class which his nearest neighbor belongs to.

In general, the classification process consisted of the following steps. Firstly, for the sample to be classified, we can obtain the linguistic label for every attribute. Secondly, matching between sample and rule about linguistic labels is conducted in rule base. If the matched rule is found, the rule's consequent part is the class label of sample. Otherwise, according to Definition 2, the sample's class label is determined by the consequent part of the nearest rule of sample.

3. Methodology

3.1. The lager ethnic groups in China

China is a united and multi-ethnic country since Qin Dynasty was founded in 221 BC. Nowadays, there are 56 ethnic groups living in the homeland. Table 1 gives several larger ones' description on native language, population, subgroups and religion. Obviously, Han Chinese is the largest ethnic group (91.51%), and other ones is often called ethnic minorities (8.49%). China's ethnic groups live together over vast areas while some live in individual concentrated communities in small areas. Fig. 1 shows the geographical distribution of these ethnic groups. Although ethnic minorities are few in population, their autonomous region area accounts for 64.2% of China's territory.

On account of the unique characteristics of China's ethnic groups, we pick several ethnic groups among them which can be distinguished with following features.

- *Population:* To facilitate face image collection, the ethnic groups with a larger population are chosen for research. Because people belonging to these ethnic groups are more likely found, which guarantees the number of face images in Chinese ethnic face database for every ethnic group.
- *Settlements:* Since the physical appearances of ethnic groups are influenced by ecological environment where they dwell in, the settlement of chosen ethnic groups must be separated in geography. Moreover, mixed marriage is more frequent in living together areas than compact community, which causes uncertain effects on appearance evolution. That is another reason for emphasizing geographic separation.

Under the guidance of above principles, we finally confirm that Korean, Uyghur and Zhuang are considered as research focus. Because, these ethnic groups respectively live in northeast, northwest, and southern of China and their population is relatively larger.

3.2. Extracting multi-ethnic facial geometric features

Landmarks are significant geometric features extracted from face, which are the basis of building component facial geometric features. Landmarks are usually categorized as mathematical, anatomical and pseudo-landmarks. Anatomical landmarks are characterized by biological meaning and strict definition by experts. Mathematical landmarks are defined on the basis of mathematical or geometric properties. Pseudo-landmarks are distinguished by locations around the outline of an object or between two existing anatomical or mathematical landmarks. For analyzing the diversity in multi-ethnic facial features, anatomical landmarks are the most important on account of their biological

Table 1
The statistics of larger ethnic groups in China [23].

Name	Native language	Population	Subgroups	Religion
Han Chinese	Chinese	1300 million	Cantonese, Chuanqing, Fuzhouese, Min, Gan, Hakka, Hunanese, Hoklo, Shanghainese, Taishanese, Tanka, Teochew	N/A
Hui	Chinese	10 million	N/A	Sunni Islam
Manchu	Chinese	10.4 million	N/A	N/A
Mongols	Mongolian	5.9 million	N/A	Tibetan Buddhism
Tibetan	Tibetan	6.2 million	N/A	Tibetan Buddhism
Uyghur	Uyghur	10 million	N/A	Sunni Islam
Zhuang	Zhuang	16.9 million	N/A	Moism
Koreans	Korean	1.8 million	N/A	N/A

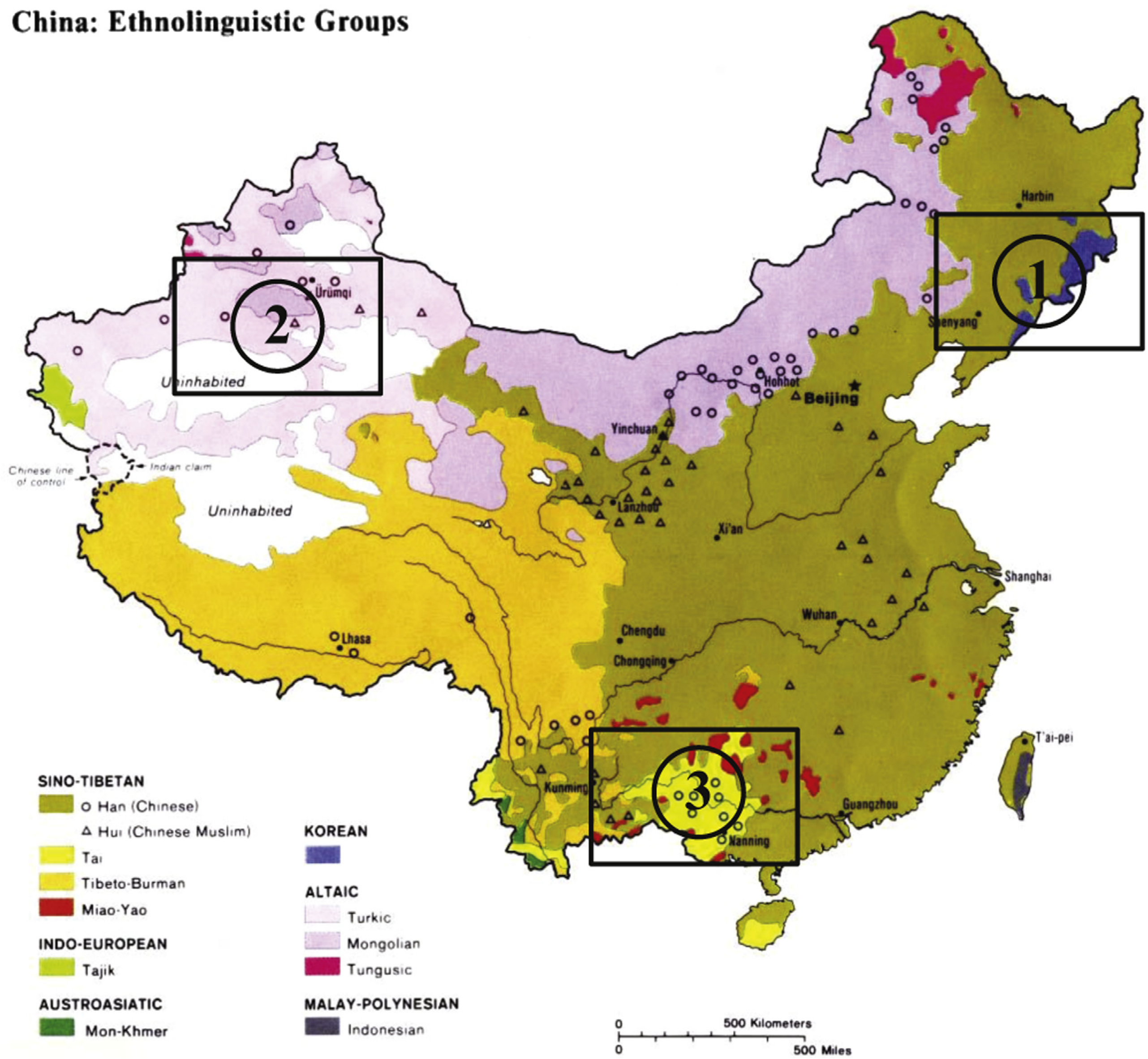

Fig. 1. The geographical distribution of larger ethnic groups in China [23] (1: Koreans, 2: Uyghur, 3: Zhuang).

implications. The other landmarks are commonly applied in automated face recognition based on geometric features.

In this paper, we adopt active shape models with SIFT Descriptors and MARS [19] as landmarking method because of its efficiency and performance against available techniques for automatic face landmarking on frontal faces. Before landmarking, face image normalization must be finished, including face detection, graying, smoothing, binarizing and scale variation. Face images from Chinese ethic face database are processed as proper binarized face images. Stasm 4.1.0 (http://www.milbo.users.sonic.net/stasm/), a C++ software library built based on [19], is introduced for finding features points in face. 77 landmarks are obtained by this way, Fig. 2(b) indicates their locations on face image. In order to describe these locations clearly, 77 landmarks are divided into five regions. They are *facial contours*, *eyebrow*, *eye*, *nose* and *mouth* distinguished by *red*, *green*, *blue*, *brown* and *yellow* points in Fig. 2(c). The implications of landmarks are displayed in Table 2. Among these landmarks, the following ones are used in constructing anthropometric measurements.

After synthesizing two kinds of landmark systems, we pick up *tr*, *n*, *ex*, *en*, *al*, *ch*, *go* and *gn* for establishing geometric features, because the coordinate of these points can be directly achieved or indirectly computed from 77 landmarks generated by STASM. The mapping relationship exists between two-dimensional face images and landmarks, so we adopt a complex number notation for every landmark [24]. The set of N landmarks is denoted as $\{l_i : l_i \in \mathbb{C}\}_{i=1}^{N}$. Then, the face F can be modeled by a N-dimensional vector as follows:

$$F = [l_1, l_2, \ldots, l_N]^T \tag{7}$$

$$l_i = x_i + \sqrt{-1}y_i \tag{8}$$

where (x_i, y_i) is coordinate of the landmark l_i.

The difference in scale between face images is almost insurmountable without artificial cutting and alignment. Therefore, the distances between landmarks are not directly viewed as linear measurements for anthropometric analysis. In this paper, ratios are employed as features instead of distance and defined for the Euclidean distance among landmarks, because they are immune to translation,

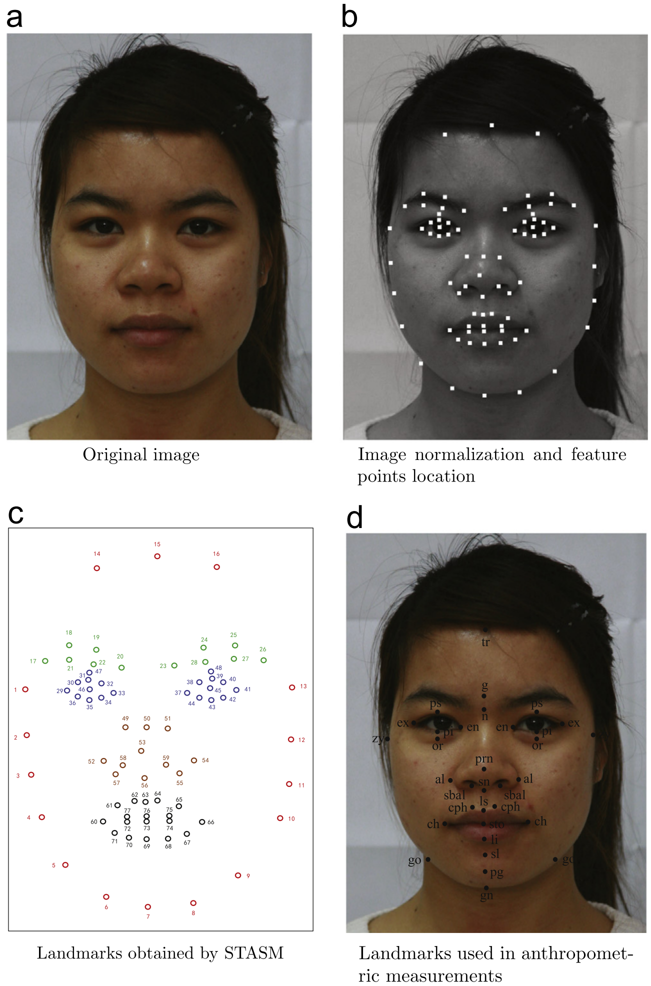

Fig. 2. Landmarks used in this work. (a) Original image. (b) Image normalization and feature points location. (c) Landmarks obtained by STASM. (d) Landmarks used in anthropometric measurements. (For interpretation of the references to color in the text, the reader is referred to the web version of this paper.)

Table 2
The implications of feature points.

Number	Name	Implication
1	zygion (zy)	Most lateral point of each zygomatic arch
5 or 9	gonion (go)	Most lateral point on the mandibular angle
7	gnathion (gn)	Lowest median landmark on the lower border of the mandible
15	tragion (tr)	Notch on upper margin of the tragus
29 or 41	exocanthion (ex)	Point at the outer commissure of the eye fissure
33 or 37	endocanthion (en)	Point at the inner commissure of the eye fissure
52 or 54	alare (al)	Most lateral point of each alar contour
56	subnasale (sn)	Midpoint of the angle at the columella base where the lower border of the nasal septum and the surface of the upper lip meet
60 or 66	cheilion (ch)	Point located at each labial commissure
0.5*(20+23)	nasion (n)	Point in the midline of both the nasal root and nasofrontal suture

scaling and two-dimensional rotation. The defined calculation method of ratios requires up to four landmarks. For the convenience of computing, we figure up the ratios relative to one benchmark distance rather than the ratios between two arbitrary landmarks. The benchmark distance is assigned by inter-eye distance. Its length is appropriate, which enable that normalized distances are insensitive to noise. Moreover, the two landmarks used in constructing inter-eye distance are more precisely located than other landmarks.

By this means, we can calculate the Euclidean distance between landmarks as follows:

$$d(l_i, l_j) = \|l_i - l_j\|_2. \tag{9}$$

If we let l_a and l_b are the left and right eye pupil location, respectively, benchmark distance can be defined as $d(l_a, l_b)$. The normalized distance for a face are then denoted as

$$r(l_i, l_j) = \frac{d(l_i, l_j)}{d(l_a, l_b)}, \quad \forall l_i, l_j \in \{l_1, \ldots, l_N\} \tag{10}$$

For a given face F, there is a corresponding set of N ratios $\{r_i : ri \in \mathbb{R}\}_{i=1}^{N}$. The face F can be represented by a N-dimensional vector of ratios as

$$F = [r_1, r_2, \ldots, r_N]^T. \tag{11}$$

Finally, we choose 12 geometric features based on 14 landmarks. Their normalized distances are computed by (10). The implications of these features are shown in Fig. 3.

3.3. Generating linguistic fuzzy rules

From the viewpoint of human in ethnic group recognition, the result directly depends on the information obtained from face and knowledge stored in memory. The knowledge is learned from ethnic group recognition experiences. Essentially, the knowledge is denoted as rule base in practice. For example, almost all interviewers give their reasons with rule like “Because the nose is very wide, the forehead is high, and so on.” when they are required to recognize a human's ethnic attribute and explain the justification. So, we use fuzzy concept to describe ethnic facial features rather than accurate measurements. That's why we introduce fuzzy rule-based systems (WM method) to perform semantic description for face features of larger Chinese ethnic groups.

In addition, the operation of extracting multi-ethnic facial geometric features is influenced by uncertainty. The uncertainty is embodied in imprecise landmarks. On the one hand, due to the face image quality or other uncertain factors, landmarks obtained by STASM may be influenced in location accuracy. On the other hand, the substantial differences in morphology between ethnic groups lie in craniofacial features. The Chinese ethnic face database consisted of front face image. There is uncertainty associated with mapping 2D landmarks in face surface to 3D ones in craniofacial bone. The details of the uncertainty are given in [25,26]. The WM method improved in this paper is an effective tool to process uncertain data.

In conclusion, the implementation steps for the proposed method are displayed as follows:

Step 1: *Definition of membership functions.* In this paper, we have 12 input variables. They are x_1 (*face width*), x_2 (*mandible*

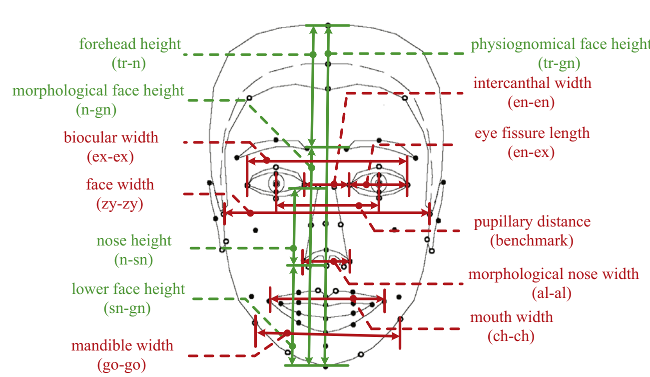

Fig. 3. The definition of geometric features.

width), x_3 (*physiognomical face height*), x_4 (*morphological face height*), x_5 (*forehead height*), x_6 (*lower face height*), x_7 (*morphological nose width*), x_8 (*mouth width*), x_9 (*biocular width*), x_{10} (*eye fissure length*), x_{11} (*nose height*) and x_{12} (*intercanthal width*). *Korean* (y_1), *Uyghur*(y_2) and *Zhuang*(y_3) are viewed as output variables. Let $d_c = c_i^{max} - c_i^{min}$, where c_i^{max} and c_i^{min} are the maximal and minimal value of *i*th attribute, respectively. The triangular membership functions are used for input variables' fuzzy division except for the two end membership functions, which are in trapezoid forms. The following five parameter sets specify the five membership function for input variable c_i: $[c_i^{min}, c_i^{min}, c_i^{min}, c_i^{min}+0.25d_c]$, $[c_i^{min}, c_i^{min}+0.25d_c, c_i^{min}+0.5d_c]$, $[c_i^{min}+0.25d_c, c_i^{min}+0.5d_c, c_i^{min}+0.75d_c]$, $[c_i^{min}+0.5d_c, c_i^{min}+0.75d_c, c_i^{max}]$ and $[c_i^{min}+0.75d_c, c_i^{max}, c_i^{max}, c_i^{max}]$. Fig. 4 shows the division method.

Step 2: *Constructing sample data.* Firstly, landmarks of every facial image are obtained by STASM and denoted in the form of (8). Secondly, according to (9) and (10), corresponding ratios for features in Fig. 3 are calculated. The class label and ratios are combined into raw data as follows:

$$(r_1^p, r_2^{(p)}, \ldots, r_n^{(p)}, y^{(p)}) \tag{12}$$

where $y \in \{Korean, Uyghur, Zhuang\}$ and r_i is the *i*th geometric feature's ratio of *p*th sample.

Step 3: *Extracting fuzz rules.* A simple dataset consisting of 10 samples (see Table 3) is introduced to explicitly illustrate the method. We suppose that $A = \{A_{l_1}^{r_i}, \ldots, A_{l_5}^{r_i}\}$ and $B = \{B_{l_1}^y, B_{l_2}^y, B_{l_3}^y\}$ are the sets of linguistic labels for attribute r_i and y, respectively. For every sample $r^{(p)}$ in Table 3, the membership degree of attribute r_i is calculated by $\mu_{A^{r_i}}(r_i^{(p)})$ defined in Fig. 4. For showing the details conveniently, we define 3 fuzzy sets for every attribute in this example. By fuzzification, the raw sample dataset is converted to fuzzified one. Tables 4–6 give the membership degree of all attributes in their fuzzy sets. Then, we can construct the antecedent part with attributes' linguistic labels based on the principle of maximum membership degree, and the consequent part is assigned with sample's class label. Meanwhile, the rule weight R_W ($r^{(p)}$) for every sample $r^{(p)}$ can be computed by (3). The obtained fuzzy rule base is presented as follows.

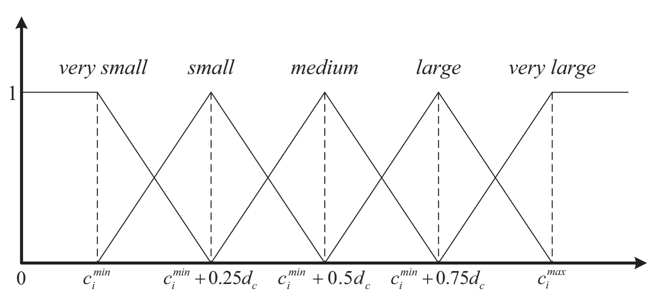

Fig. 4. The membership function of attribute c_i.

R_1 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang and $R_W^1 = 0.1950$.

R_2 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang and $R_W^2 = 0.0440$.

R_3 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang and $R_W^3 = 0.0190$.

R_4 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean and $R_W^4 = 0.0246$.

R_5 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean and $R_W^5 = 0.1609$.

R_6 **IF** r_1 is $A_{l_1}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_1}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_1}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang and $R_W^6 = 0.0147$.

R_7 **IF** r_1 is $A_{l_1}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_1}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_1}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang and $R_W^7 = 0.0223$.

R_8 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_3}^{r_3}$, r_4 is $A_{l_3}^{r_4}$, r_5 is $A_{l_1}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_3}^{r_1}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean and $R_W^8 = 0.0245$.

R_9 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_3}^{r_3}$, r_4 is $A_{l_3}^{r_4}$, r_5 is $A_{l_1}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_3}^{r_1}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean and $R_W^9 = 0.0472$.

R_{10} **IF** r_1 is $A_{l_3}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_1}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean and $R_W^{10} = 0.0587$.

Step 4: *Refining the rule base.* At this stage, the following cases are considered to simplify the rule base $R_B = \{R_1, R_2, \ldots, R_n\}$. For given $R_i, R_j \in R_B$, if their antecedent and consequent parts are identical, the rule with less weight will be deleted from rule base. Moreover, if R_i and R_j are same except for the consequent part, their support degrees are calculated by (4), and we only reserve the rule with larger support degree. The above mentioned rule base is refined as the following one.

R_1 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang, $R_W^1 = 0.1950$, $R_S^1 = 0.0258$.

R_7 **IF** r_1 is $A_{l_1}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_1}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_1}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Zhuang, $R_W^7 = 0.0223$ and $R_S^7 = 0.0017$.

Table 3
Raw sample data.

r	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	y
$r^{(1)}$	2.1792	1.4375	2.9745	2.1153	0.8599	1.0995	0.6536	0.7326	1.5679	0.3761	1.0159	0.6336	Zhuang
$r^{(2)}$	2.1576	1.2983	2.9948	2.1863	0.8111	1.1779	0.6491	0.7398	1.5485	0.3699	1.0085	0.6301	Zhuang
$r^{(3)}$	2.1011	1.3789	2.8915	2.0478	0.8438	1.0187	0.7107	0.8340	1.5075	0.3602	1.0290	0.6393	Zhuang
$r^{(4)}$	2.1442	1.2678	2.9007	2.0595	0.8415	1.0140	0.7145	0.7863	1.4930	0.3681	1.0461	0.6331	Korean
$r^{(5)}$	2.2603	1.3958	3.0207	2.1511	0.8699	1.1668	0.6563	0.8020	1.5468	0.3854	0.9843	0.6251	Korean
$r^{(6)}$	2.0305	1.4120	2.8624	1.9265	0.9359	1.0705	0.6602	0.7205	1.4497	0.3605	0.8560	0.6402	Zhuang
$r^{(7)}$	2.0679	1.4603	2.8517	1.9502	0.9017	1.0682	0.6567	0.7352	1.4416	0.3729	0.8821	0.6371	Zhuang
$r^{(8)}$	2.1552	1.3377	3.0756	2.3187	0.7606	1.1881	0.6782	0.7486	1.5918	0.3732	1.1307	0.6264	Korean
$r^{(9)}$	2.2158	1.4065	3.0928	2.3173	0.7756	1.1065	0.6329	0.7525	1.5984	0.3653	1.2108	0.6336	Korean
$r^{(10)}$	2.4058	1.3859	2.8614	2.0614	0.8028	1.0402	0.6831	0.7625	1.4702	0.3763	1.0213	0.6237	Korean

Table 4
Fuzzified sample data (Part 1).

r	r_1			r_2			r_3			r_4		
	$A_{l_1}^{r_1}$	$A_{l_2}^{r_1}$	$A_{l_3}^{r_1}$	$A_{l_1}^{r_2}$	$A_{l_2}^{r_2}$	$A_{l_3}^{r_2}$	$A_{l_1}^{r_3}$	$A_{l_2}^{r_3}$	$A_{l_3}^{r_3}$	$A_{l_1}^{r_4}$	$A_{l_2}^{r_4}$	$A_{l_3}^{r_4}$
$r^{(1)}$	0.0792	**0.9208**	0.0000	0.0000	**0.8865**	0.1135	0.0000	**0.8931**	0.1069	0.0000	**0.8815**	0.1185
$r^{(2)}$	0.1623	**0.8377**	0.0000	0.3501	**0.6499**	0.0000	0.0000	**0.8096**	0.1904	0.0000	**0.6639**	0.3361
$r^{(3)}$	0.3793	**0.6207**	0.0000	0.0819	**0.9181**	0.0000	0.2352	**0.7648**	0.0000	0.0883	**0.9117**	0.0000
$r^{(4)}$	0.2134	**0.7861**	0.0000	0.4517	**0.5483**	0.0000	0.1973	**0.8027**	0.0000	0.0524	**0.9476**	0.0000
$r^{(5)}$	0.0000	**0.7677**	0.2323	0.0255	**0.9745**	0.0000	0.0000	**0.7028**	0.2972	0.0000	**0.7719**	0.2280
$r^{(6)}$	**0.6508**	0.3492	0.0000	0.0000	**0.9716**	0.0284	0.3550	**0.6450**	0.0000	0.4599	**0.5401**	0.0000
$r^{(7)}$	**0.5068**	0.4932	0.0000	0.0000	**0.8108**	0.1892	0.3990	**0.6010**	0.0000	0.3873	**0.6127**	0.0000
$r^{(8)}$	0.1714	**0.8286**	0.0000	0.2190	**0.7810**	0.0000	0.0000	0.4764	**0.5236**	0.0000	0.2584	**0.7416**
$r^{(9)}$	0.0000	**0.9385**	0.0615	0.0000	**0.9897**	0.0103	0.0000	0.4053	**0.5947**	0.0000	0.2627	**0.7373**
$r^{(10)}$	0.0000	0.2085	**0.7915**	0.0585	**0.9415**	0.0000	0.3589	**0.6411**	0.0000	0.0467	**0.9533**	0.0000

Table 5
Fuzzified sample data (Part 2).

r	r_5			r_6			r_7			r_8		
	$A_{l_1}^{r_5}$	$A_{l_2}^{r_5}$	$A_{l_3}^{r_5}$	$A_{l_1}^{r_6}$	$A_{l_2}^{r_6}$	$A_{l_3}^{r_6}$	$A_{l_1}^{r_7}$	$A_{l_2}^{r_7}$	$A_{l_3}^{r_7}$	$A_{l_1}^{r_8}$	$A_{l_2}^{r_8}$	$A_{l_3}^{r_8}$
$r^{(1)}$	0.1045	**0.8955**	0.0000	0.0000	**0.9479**	0.0521	0.1160	**0.8840**	0.0000	0.2497	**0.7503**	0.0000
$r^{(2)}$	0.4400	**0.5600**	0.0000	0.0000	**0.6172**	0.3828	0.1567	**0.8433**	0.0000	0.2082	**0.7918**	0.0000
$r^{(3)}$	0.2156	**0.7844**	0.0000	0.2887	**0.7113**	0.0000	0.0000	**0.5996**	0.4004	0.0000	**0.6611**	0.3389
$r^{(4)}$	0.2312	**0.7688**	0.0000	0.3085	**0.6915**	0.0000	0.0000	**0.5645**	0.4355	0.0000	**0.9378**	0.0622
$r^{(5)}$	0.0360	**0.9640**	0.0000	0.0000	**0.6642**	0.3358	0.0919	**0.9081**	0.0000	0.0000	**0.8468**	0.1532
$r^{(6)}$	0.000	**0.5820**	0.4180	0.0701	**0.9299**	0.0000	0.0568	**0.9432**	0.0000	0.3200	**0.6800**	0.0000
$r^{(7)}$	0.0000	**0.8172**	0.1828	0.0801	**0.9199**	0.0000	0.0887	**0.9113**	0.0000	0.2344	**0.7656**	0.0000
$r^{(8)}$	**0.7874**	0.2126	0.0000	0.0000	**0.5744**	0.4256	0.0000	**0.8932**	0.1068	0.1571	**0.8429**	0.0000
$r^{(9)}$	**0.6846**	0.3154	0.0000	0.0000	**0.9182**	0.0818	0.3037	**0.6963**	0.0000	0.1345	**0.8655**	0.0000
$r^{(10)}$	0.4973	**0.5027**	0.0000	0.1983	**0.8017**	0.0000	0.0000	**0.8491**	0.1509	0.0763	**0.9237**	0.0000

Table 6
Fuzzified sample data (Part 3).

r	r_9			r_{10}			r_{11}			r_{12}		
	$A_{l_1}^{r_9}$	$A_{l_2}^{r_9}$	$A_{l_3}^{r_9}$	$A_{l_1}^{r_{10}}$	$A_{l_2}^{r_{10}}$	$A_{l_3}^{r_{10}}$	$A_{l_1}^{r_{11}}$	$A_{l_2}^{r_{11}}$	$A_{l_3}^{r_{11}}$	$A_{l_1}^{r_{12}}$	$A_{l_2}^{r_{12}}$	$A_{l_3}^{r_{12}}$
$r^{(1)}$	0.0000	**0.8227**	0.1773	0.1614	**0.8386**	0.0000	0.0000	**0.9445**	0.0555	0.0000	**0.8273**	0.1727
$r^{(2)}$	0.0000	**0.9797**	0.0203	0.2758	**0.7242**	0.0000	0.0000	**0.9801**	0.0199	0.0000	**0.9366**	0.0634
$r^{(3)}$	0.3112	**0.6888**	0.0000	0.4504	**0.5496**	0.0000	0.0000	**0.8804**	0.1196	0.0000	**0.6492**	0.3508
$r^{(4)}$	0.4282	**0.5718**	0.0000	0.3074	**0.6926**	0.0000	0.0000	**0.7981**	0.2019	0.0000	**0.8420**	0.1579
$r^{(5)}$	0.0000	**0.9937**	0.0063	0.0000	**0.9934**	0.0066	0.0972	**0.9028**	0.0000	0.0963	**0.9037**	0.0000
$r^{(6)}$	**0.7787**	0.2213	0.0000	0.4468	**0.5532**	0.0000	**0.7189**	0.2811	0.0000	0.0000	**0.6209**	0.3791
$r^{(7)}$	**0.8444**	0.1556	0.0000	0.2507	**0.7793**	0.0000	**0.5926**	0.4074	0.0000	0.0000	**0.7192**	0.2808
$r^{(8)}$	0.0000	**0.6297**	0.3703	0.2155	**0.7845**	0.0000	0.0000	0.3881	**0.6119**	0.0535	**0.9465**	0.0000
$r^{(9)}$	0.0000	**0.5765**	0.4235	0.3586	**0.6414**	0.0000	0.0000	0.0000	**1.0000**	0.0000	**0.8275**	0.1725
$r^{(10)}$	**0.6130**	0.3870	0.0000	0.1590	**0.8410**	0.0000	0.0000	**0.9179**	0.0821	0.1383	**0.8617**	0.0000

R_9 **IF** r_1 is $A_{l_2}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_3}^{r_3}$, r_4 is $A_{l_3}^{r_4}$, r_5 is $A_{l_1}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_2}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_3}^{r_1}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean, $R_W^9 = 0.0472$ and $R_S^9 = 0.0072$.

R_{10} **IF** r_1 is $A_{l_3}^{r_1}$, r_2 is $A_{l_2}^{r_2}$, r_3 is $A_{l_2}^{r_3}$, r_4 is $A_{l_2}^{r_4}$, r_5 is $A_{l_2}^{r_5}$, r_6 is $A_{l_2}^{r_6}$, r_7 is $A_{l_2}^{r_7}$, r_8 is $A_{l_2}^{r_8}$, r_9 is $A_{l_1}^{r_9}$, r_{10} is $A_{l_2}^{r_{10}}$, r_{11} is $A_{l_2}^{r_{11}}$, r_{12} is $A_{l_2}^{r_{12}}$ **THEN** y is Korean, $R_W^{10} = 0.0587$ and $R_S^{10} = 0.0059$.

In practical application, we expect a rule base without redundancy and conflict. Rule weight is used for removing redundant rules, and support degree is adopted to keep rule base's consistency. Finally, we can select the rules with larger support degree for describing the ethnic facial characteristics, because support degree reflects the capability of fuzzy rule in representing samples' inherent regularity.

3.4. Application in learning ethnic groups from face

The rule base generated by method proposed in Section 3.3 effectively depicts multi-ethnic facial features. So, we intend to utilize these rules for learning ethnicity from face. In order to describe facial features fully, we define perimeter and area measurements for eyebrow, eye, nose and mouth except for line measurements proposed in Section 3.3. The calculation method of these features is displayed as follows.

- *Eyebrow perimeter:* $f_{13}=\frac{1}{2}[r(l_{16},l_{17})+r(l_{17},l_{18})+r(l_{18},l_{19})+r(l_{19},l_{20})+r(l_{20},l_{21})+r(l_{22},l_{23})+r(l_{23},l_{24})+r(l_{24},l_{25})+r(l_{25},l_{26})+r(l_{26},l_{27})+r(l_{16},l_{17})+r(l_{27},l_{22})]$.
- *Eye perimeter:* $f_{14}=\frac{1}{2}[r(l_{30},l_{31})+r(l_{31},l_{32})+r(l_{32},l_{33})+r(l_{33},l_{34})+r(l_{34},l_{35})+r(l_{35},l_{36})+r(l_{36},l_{37})+r(l_{37},l_{30})+r(l_{40},l_{41})+r(l_{41},l_{42})+r(l_{42},l_{43})+r(l_{43},l_{44})+r(l_{44},l_{45})+r(l_{45},l_{46})+r(l_{46},l_{47})+r(l_{47},l_{40})]$.
- *Nose perimeter:* $f_{15}=r(l_{48},l_{49})+r(l_{49},l_{50})+r(l_{50},l_{58})+r(l_{58},l_{57})+r(l_{57},l_{56})+r(l_{56},l_{55})+r(l_{55},l_{54})+r(l_{54},l_{48})$;
- *Mouth perimeter:* $f_{16}=r(l_{59},l_{60})+r(l_{60},l_{61})+r(l_{61},l_{62})+r(l_{62},l_{63})+r(l_{63},l_{64})+r(l_{64},l_{65})+r(l_{65},l_{72})+r(l_{72},l_{73})+r(l_{73},l_{74})+r(l_{74},l_{75})+r(l_{75},l_{76})+r(l_{76},l_{59})$.
- *Eyebrow area:* $f_{17}=\frac{1}{4}[(x_{18}y_{19}-x_{19}y_{18})+(x_{19}y_{20}-x_{20}y_{19})+(x_{20}y_{21}-x_{21}y_{20})+(x_{21}y_{16}-x_{16}y_{21})+(x_{16}y_{17}-x_{17}y_{16})+(x_{17}y_{18}-x_{18}y_{17})+(x_{22}y_{27}-x_{27}y_{22})+(x_{27}y_{26}-x_{26}y_{27})+(x_{26}y_{25}-x_{25}y_{26})+(x_{25}y_{24}-x_{24}y_{25})+(x_{24}y_{23}-x_{23}y_{24})+(x_{23}y_{22}-x_{22}y_{23})]$.
- *Eye area:* $f_{18}=\frac{1}{4}[(x_{34}y_{35}-x_{35}y_{34})+(x_{35}y_{36}-x_{36}y_{35})+(x_{36}y_{37}-x_{37}y_{36})+(x_{37}y_{30}-x_{30}y_{37})+(x_{30}y_{31}-x_{31}y_{30})+(x_{31}y_{32}-x_{32}y_{31})+(x_{32}y_{33}-x_{33}y_{32})+(x_{33}y_{34}-x_{34}y_{33})+(x_{40}y_{47}-x_{47}y_{40})+(x_{47}y_{46}-x_{46}y_{47})+(x_{46}y_{45}-x_{45}y_{46})+(x_{45}y_{44}-x_{44}y_{45})+(x_{44}y_{43}-x_{43}y_{44})+(x_{43}y_{42}-x_{42}y_{43})+(x_{42}y_{41}-x_{41}y_{42})+(x_{41}y_{40}-x_{40}y_{41})]$.
- *Nose area:* $f_{19}=\frac{1}{2}[(x_{54}y_{48}-x_{48}y_{54})+(x_{48}y_{49}-x_{49}y_{48})+(x_{49}y_{50}-x_{50}y_{49})+(x_{50}y_{58}-x_{58}y_{50})+(x_{58}y_{57}-x_{57}y_{58})+(x_{57}y_{56}-x_{56}y_{57})+(x_{56}y_{55}-x_{55}y_{56})+(x_{55}y_{54}-x_{54}y_{55})]$.
- *Mouth area:* $f_{20}=\frac{1}{2}[(x_{65}y_{64}-x_{64}y_{65})+(x_{64}y_{63}-x_{63}y_{64})+(x_{63}y_{62}-x_{62}y_{63})+(x_{62}y_{61}-x_{61}y_{62})+(x_{61}y_{60}-x_{60}y_{61})+(x_{60}y_{59}-x_{59}y_{60})+(x_{59}y_{76}-x_{76}y_{59})+(x_{76}y_{75}-x_{75}y_{76})+(x_{75}y_{74}-x_{74}y_{75})+(x_{74}y_{73}-x_{73}y_{74})+(x_{73}y_{72}-x_{72}y_{73})+(x_{72}y_{65}-x_{65}y_{72})]$.

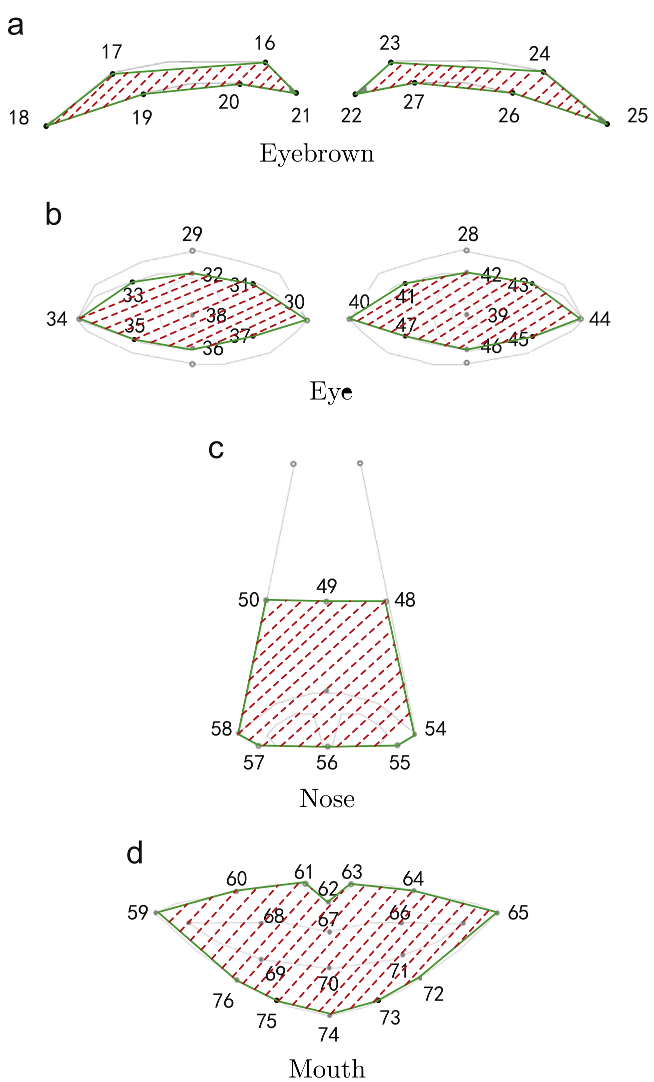

Fig. 5. Added geometric features. (a) Eyebrown. (b) Eye. (c) Nose. (d) Mouth.

where (x_i, y_i) is the coordinate of ith landmark (Fig. 5).

In this work, 20 features are extracted from landmarks. The redundancy and dependency of features may jeopardize the classification accuracy. Meanwhile, the algorithm complexity of fuzzy rule extraction depends on the number of features. Inordinate number of features tends to cause the phenomenon of "rule explosion". Therefore, we should select the optimized subset from feature set. MRMR (Min-Redundancy and Max-Relevance) [27] is introduced to realize feature selection in this paper, whose intention is that m features with max-relevance and min-redundancy for target class are selected from feature space. The definition of max-relevance and min-redundancy are shown as follows:

$$\max D(S,c), D=\frac{1}{|S|}\sum_{x_i\in S} I(x_i;c); \tag{13}$$

$$\min R(S), R=\frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i;x_j); \tag{14}$$

where S is the set of features, c is the objective class, $I(x_i;c)$ is the mutual information between feature i and target class c and $I(x_i,x_j)$ is the mutual information between feature x_i and x_j. For given random variables x and y, we can calculate the mutual information between them with:

$$I(x;y)=\int\int p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\,dx\,dy. \tag{15}$$

where $p(x)$ or $p(y)$ is the probability distribution of variable x or y and $p(x,y)$ is the joint probability distribution of variables x and y. The mutual information between multivariate S_m and target class c is defined as:

$$I(S_m;c)=\int\int p(S_m,c)\log\frac{p(S_m,c)}{p(S_m)p(c)}\,dS_m\,dc. \tag{16}$$

Then, we can select features by criterion given as follows:

$$\max \Phi(D,R),\quad \Phi=D-R. \tag{17}$$

The details on mRMR can be found in [27].

More formally, the procedures are shown in Fig. 6 and stated as follows: Let X be a set of samples in training data set and $F=\{f_1,f_2,\ldots,f_n\}$ is the feature set of X. For each $f_i\in F$, we find the maximum value f_i^{max} and minimum value f_i^{min}. The interval

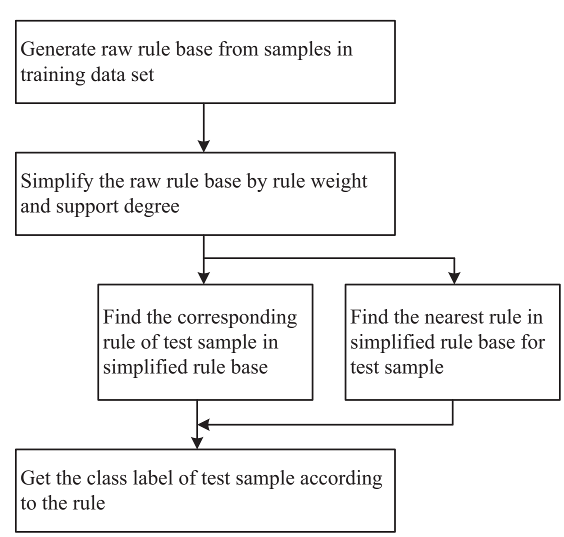

Fig. 6. The flow chart of ethnicity recognition algorithm.

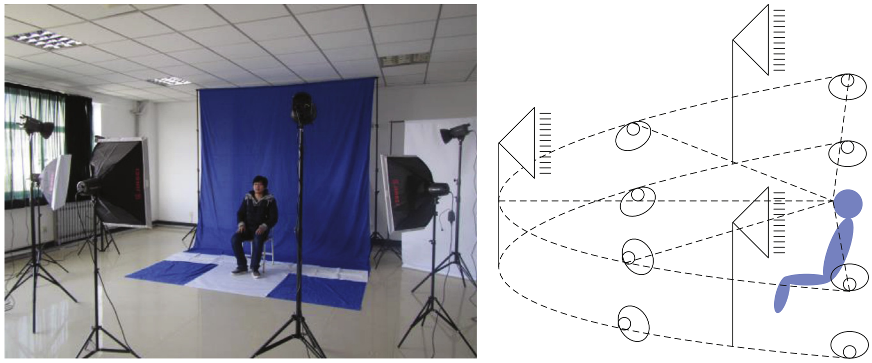

Fig. 7. The acquisition environment of Chinese ethnic groups DB.

Zhuang Uyghur Korean

Fig. 8. The sample images of Chinese ethnic DB.

consisting by f_i^{max} and f_i^{min} is equivalently divided into $k-1$ parts. Then we can get the parameter set $P_i=\{f_i^{min}, f_i^{min}+((j-1)/(k-1))(f_i^{max}-f_i^{min}), \ldots, f_i^{max}\}$ of feature f_i. For each feature f_i, the fuzzy set $F_i=\{\mu_{A_{l_1}}^{f_i}, \mu_{A_{l_2}}^{f_i}, \ldots, \mu_{A_{l_k}}^{f_i}\}$ is constructed for building membership function. Fig. 4 indicates the case when $k=5$. For each sample $x_m \in X$, the membership degree is calculated in every feature f_i. Afterwards, the linguistic label is assigned with the principle of maximal membership degree, $l_m^i = \mathrm{argmax}_{A_{l_j}, 1 \le j \le k}\{\mu_{A_{l_j}}^{f_i}(x_m)\}$. The rule weight $R_W{}^m$ for each sample x_m can be obtained by (3). Finally, the corresponding rule R^m of every sample x_m is generated, whose antecedent part consisted of linguistic labels and consequent part is assigned by class label of sample x_m.

After above procedure, the raw rules are extracted from samples in training data set. In the next moment, simplification is carried out for keeping the consistency and reducing the redundancy of rule base. Let R_B be the rule base built in last step, for each rule R^m in rule base R_B, we can classify R_m into category K_i if the antecedent and consequent parts are identical with the rules' belonging to R_B. For every class K_i, the rule with maximum rule weight is viewed as the center of rule category K_i and denoted as $R_{K_i}^c$. For any rules $R_i, R_j \in R_B$, if R_i and R_j are identical except for their rule category $(K_i \ne K_j)$, we get support degree for $R_{K_i}^c$ and $R_{K_j}^c$ by (4). Then, the rules belonging to the class with less support degree are removed. For each rule class $K_i \in K$, all rules are removed from rule base R_B except for the rule center $R_{K_i}^c$, which finally forms the simplified rule base R_S.

Given a test instance x_t, steps $6 \rightarrow 12$ are executed to get the antecedent part of fuzzy rule R_t for instance x_t. For each rule $R^m \in R_S$, the consequent part of rule R^m is viewed as class label of instance x_t and returned if the antecedent part of R^m is same with R_t. If none rule's antecedent part matches with R_t in simplified rule base R_S, we need to find the nearest rule for classifying the instance x_t. For each $K_i \in K$, the Euclidean distance between x_t and $x_c{}^i$, where $x_c{}^i$ is the corresponding sample of rule center $R_{K_i}^c$. The consequent part of $R_{K_q}^c$ is returned if the Euclidean distance between x_t and $x_c{}^q$ is the shortest.

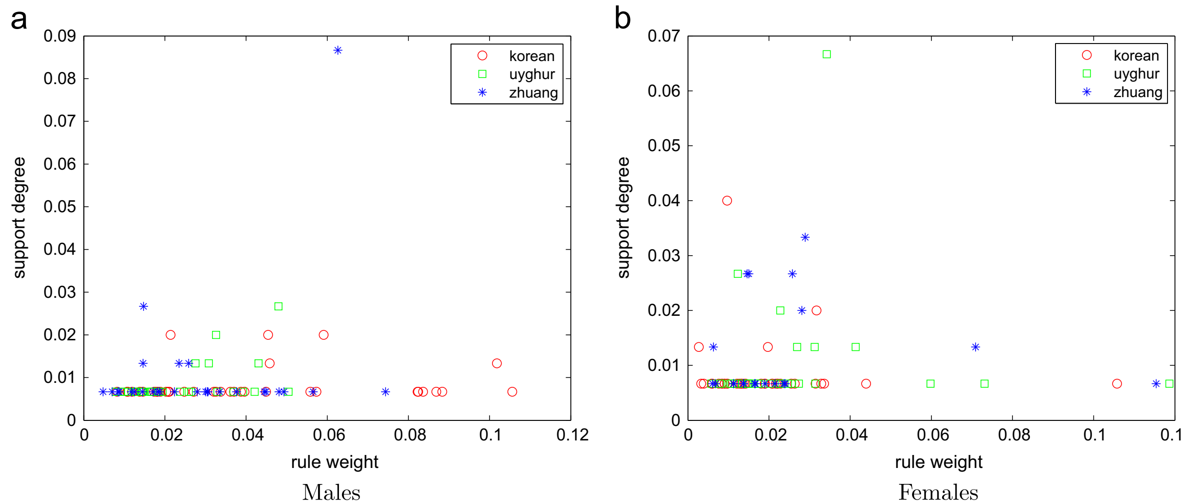

Fig. 9. The statistics of simplified rules. (a) Males. (b) Females.

Algorithm 1. Ethnicity recognition.

Input: The training data set, X; The number of fuzzy sets in every feature, k; The test sample, x_t.
Output: The class label of test sample x_t, C_t;
1: **for** each feature f_i in X **do**
2: find the maximum value f_i^{max} and minimum value f_i^{min};
3: compute the parameter set $P_i = \left\{f_i^{min}, f_i^{min} + \frac{j-1}{k-1}(f_i^{max} - f_i^{min}), \ldots, f_i^{max}\right\}$, $j = 2$,...., $k-1$;
4: construct the fuzzy set $F_i = \{\mu_{A_{l_1}}^{f_i}, \mu_{A_{l_2}}^{f_i}$,...., $\mu_{A_{l_k}}^{f_i}\}$ and build triangular membership function (see Fig. 4) with P_i;
5: **end for**
6: **for** each sample x_m in X **do**
7: **for** each feature f_i of sample x_m **do**
8: calculate the membership degree and get the linguistic label, $l_m^i = \text{argmax}_{A_{l_j}, 1 \le j \le k}\{\mu_{A_{l_j}}^{f_i}(x_m)\}$;
9: **end for**
10: obtain the rule weight $R_W{}^m$ of sample x_m by 3;
11: generate the rule R^m, the antecedent part of R^m is consisted of linguistic labels obtained in Step . 8, and the consequent part is assigned by class label of sample x_m;
12: **end for**
13: **for** each R^m in rule base **do**
14: the rules with identical antecedent and consequent part are classified into a category K_i, and the rule with maximum rule weight is viewed as the rule category's center $R_{K_i}^c$ for every class K_i;
15: **end for**
16: **for** each R^m in rule base **do**
17: **if** the antecedent of R^i and R^j is same, but $K_i \ne K_j$, where K_i and K_j are the class of R_i and R_j **then**
18: get support degree for $R_{K_i}^c$ and $R_{K_j}^c$ by 4, and delete the rules belong to the class with less support degree;
19: **end if**
20: **end for**
21: **for** each K_i in rule base **do**
22: remove rules except for the rule center $R_{K_i}^c$;
23: **end for**
24: execute Step. 6 ⇒ Step. 12 and get the antecedent part of fuzzy rule R_t for sample x_t;
25: **for** each R^m in simplified rule base **do**
26: **if** the antecedent part of R^m is same with R_t **then**
27: **return** C_t, the consequent part of R^m.
28: **end if**
29: **end for**
30: **if** none rule's antecedent matches with R_t in simplified rule base **then**
31: $q = \arg\min_{1 \le i \le n}\{EuclideanDistance(x_c^i, x_t)\}$ to get the nearest rule, where $x_c{}^i$ is the corresponding sample of rule $R_{K_i}^c$ and n is the number of rule category;s
32: **return** C_t, the consequent part of $R_{K_q}^c$;
33: **end if**

4. Experiments and analysis

4.1. CEFD

With reference to the method proposed in [28], we have built the Chinese ethnic face database (CEFD) including Korean, Uyghur and Zhuang. The participants are young volunteers recruited at Dalian Nationalities University. Fig. 7 shows the environment for acquiring facial images. Several tentative researches [29,30] are carried out based on it. For every ethnic group in CEFD, there are 50 people's frontal facial images, half man half women. Fig. 8 indicates sample images in CEFD. Since the sizes of facial images vary in scale, we scaled all images according to a fixed Euclidean distance between pupil centers. The landmarks of preprocessed face images are extracted by STASM. Geometric features are calculated based on landmarks. Finally, a face image is converted to a instance consisting of geometric features in term of numeric values. Then, the face database is translated to a geometric features database.

4.2. Generation and analysis of semantic description

Owing to the physiological differences between males and females, the research is divided into two parts to eliminate the disturbances caused by gender. For each feature f_i, the fuzzy linguistic labels “ *large*”, “ *medium*”, “ *small*” are assigned. The experimental results obtained by the method described in Section 3.3 are displayed as follows.

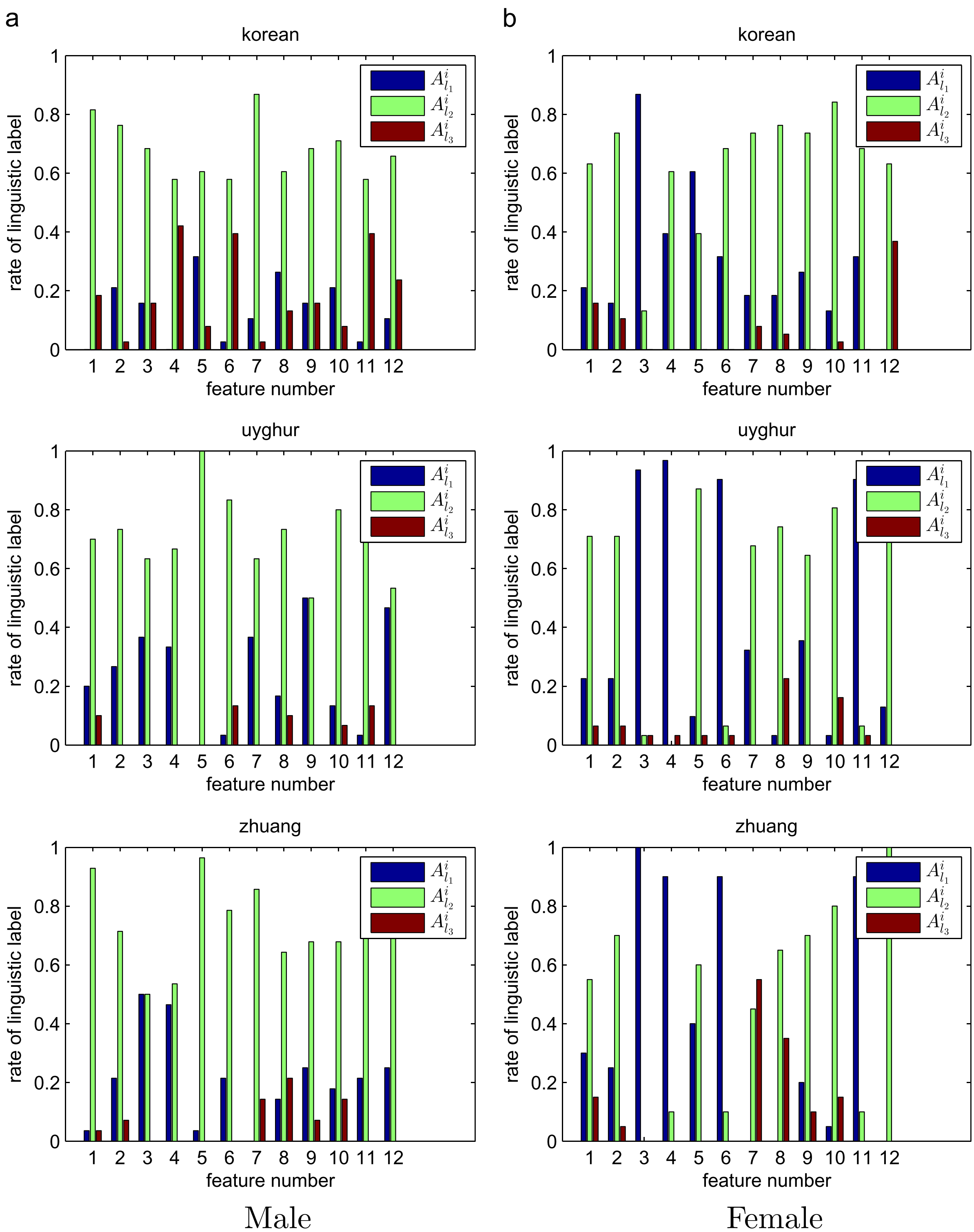

Fig. 10. The statistics of for simplified rules' linguistic labels. (a) Male. (b) Female.

For males, 150 raw rules (korean-50, uyghur-50 and zhuang-50) are extracted from geometric features dataset, and 96 rules (korean-38, uyghur-30 and zhuang-28) are remained after two simplification steps. For females, 150 raw rules are reduced to 89 rules (korean-38, uyghur-31 and 20) in the same way. Rule weight and support degree defined by (3) and (4) are applied for explaining the rule's importance in describing the ethnic facial features. Fig. 9 shows that the rules with large rule weight and high support degree are minority in the simplified rule base. However, these rules indicate the main characteristics of ethnic facial morphology.

In order to analyze the correlation between ethnicity and geometric features, we conduct a statistics about the linguistic labels' frequency of rules. Fig. 10(a) indicates that males belonging to Uyghur and Zhuang are convergent in geometric features. Because there are 5 features whose 3 linguistic labels are entirely covered by samples. Nevertheless, Korean males' features are divergent because 10 features' linguistic labels are involved.

Table 7
The score of features.

Number	Feature	Score		Order	
		Male	Female	Male	Female
1	Morphological face height	0.011	−0.041	12	16
2	Mandible width	−0.048	−0.019	19	12
3	Physiognomical face height	0.058	0.027	9	10
4	Morphological face height	0.325	0.250	1	1
5	Forehead height	0.076	0.099	6	5
6	Lower face height	−0.012	0.097	15	6
7	Morphological nose width	0.133	0.173	3	3
8	Mouth width	−0.005	0.106	14	4
9	Biocular width	0.048	0.029	10	14
10	Eye fissure length	−0.048	0.027	18	8
11	Nose height	0.074	0.030	7	11
12	Intercanthal width	0.100	0.184	4	2
13	Eyebrow area	0.106	−0.113	5	20
14	Eye area	0.071	0.089	8	7
15	Nose area	0.244	−0.041	2	15
16	Mouth area	−0.007	−0.095	16	19
17	Eyebrow perimeter	0.011	−0.055	13	17
18	Eye perimeter	−0.037	−0.077	20	18
19	Nose perimeter	0.047	0.029	11	9
20	Mouth perimeter	−0.039	−0.011	17	13

Likewise, for females, Fig. 10(b) shows that the divergence is reflected in Uyghur, and the convergence is embodied in Korean and Zhuang.

4.3. Performance comparison with other classifier in ethnicity recognition

In this work, we apply the obtained semantic description for realizing ethnicity recognition. The proposed Algorithm 1 is compared with the classic classification algorithms including Random Forest, Adaboost, Naive Bayesian, Logistic, C4.5, PART and Decision Table. Before classification, mRMR algorithm is utilized to select salient features. Table 7 gives the features' score and order. Often, according to the order, the top-k features are selected for classification.

For the sake of determining the optimal number of features, we perform experiments with 10-fold cross validation when $k = 2, 3, \ldots, 20$. The classification accuracy rate is average result of the conducted 1000 times experiments. For the purpose of investigating the capability of proposed method, we construct three subsets of the dataset including Korean, Uyghur and Zhuang (K+U+Z). They are Korean+Uyghru (K+U), Korean+Zhuang (K+Z) and Uyghur+Zhuang (U+Z). Taken together, there are 8 datasets for evaluating the algorithm performance. Fig. 11 illustrates the relation between k and accuracy rate in 8 datasets. The solid dots represent the maximum accuracy rate for every curve. In Table 8, we give the size of optimized subset and corresponding classification accuracy. The results demonstrate that using features selected by mRMR is superior to using all features in classification accuracy. The above mentioned method for ascertaining the best k is applied in Random Forest, Adaboost, Naive Bayesian, Logistic, C4.5, PART and Decision Table, the only difference is that 10-fold cross validation is executed 1 times rather than 1000 times. All the experiments are run on Weka [31] (Version 3.7) with default parameter settings, and the classification results are shown in Tables 9 and 10.

The representative methods in ensemble learning, probability-based and regression-based are used to be compared with our algorithm. The results in Table 9 show that Logistic performs outstandingly in most datasets. The differences in classification accuracy rate among Random Forest, Adaboost and our algorithm are very small, which indicates that our algorithm is competitive in classification accuracy rate. However, Random Forest, Adaboost, Naive Bayesian and Logistic

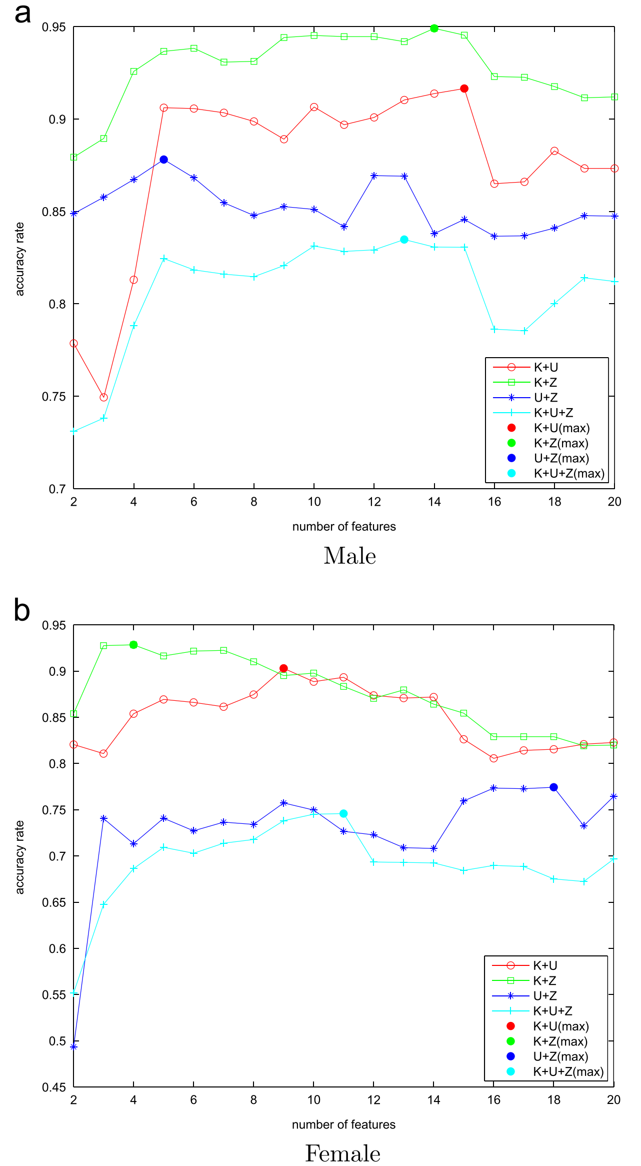

Fig. 11. The relationship between the number of features and classification accuracy rate. (a) Male. (b) Female.

Table 8
Classification accuracy of the data based on the selected features and all the features.

Data set	Selected features		All features	
	Accuracy (%)	Number	Accuracy (%)	Number
K+U (male)	91.65	15	87.33	20
K+Z (male)	94.91	14	91.20	20
U+Z (male)	87.81	5	84.75	20
K+U+Z (male)	83.48	13	81.21	20
K+U (female)	90.30	9	82.28	20
K+Z (female)	92.85	4	82.00	20
U+Z (female)	77.43	18	76.44	20
K+U+Z (female)	74.58	11	69.66	20

are poor in interpretability. Our algorithm is capable of generating linguistic interpretability for classification result.

In Table 10, rule-based classification algorithm including C4.5, PART and Decision Table are introduced to conduct comparative analysis with our algorithm. Rule-based method is characterized

Table 9
The classification result (Part 1).

Dataset	Random Forest		Adaboost		Naive Bayesian		Logistic		Our Algorithm	
	Att	Acc (%)	Att	Acc (%)	Att	Acc (%)	Att	Acc (%)	Att	Acc(%)
K+U(M)	20	94	6	94	6	95	11	**100**	15	93
K+Z(M)	8	**98**	16	97	8	**98**	9	**98**	14	94
U+Z(M)	4	92	6	91	2	90	11	**94**	14	94
K+U+Z(M)	7	86.67	2	82	6	89.33	5	**92**	13	86
K+U(F)	12	92	9	90	7	94	9	**97**	9	92
K+Z(F)	8	**97**	3	90	8	96	3	94	4	93
U+Z(F)	10	**91**	10	86	10	80	13	**91**	18	79
K+U+Z(F)	12	**87.33**	6	70	9	82.67	10	86.67	11	76

Table 10
The classification result (Part 2).

Dataset	C4.5			PART			Decision Table			Our Algorithm		
	Att	Acc (%)	Rules	Att	Acc (%)	Rules	Att	Acc (%)	Rules	Att	Acc (%)	Rules
K+U(M)	2	83	2	5	85	4	5	87	8	15	**93**	80.2
K+Z(M)	2	93	4	2	93	4	2	93	2	14	**94**	75.3
U+Z(M)	2	91	2	2	91	2	2	90	2	14	**94**	75.6
K+U+Z(M)	2	80	7	5	80	9	7	80.67	10	13	**86**	103.8
K+U(F)	4	82	5	4	83	2	4	84	4	9	**92**	58.4
K+Z(F)	3	**94**	5	13	93	4	3	93	4	4	93	26.7
U+Z(F)	10	89	7	10	**90**	3	3	77	2	18	79	78
K+U+Z(F)	10	79.33	13	9	**80**	10	10	76	12	11	76	73.8

by both remarkable classification capacity and interpretability. The experiment results illustrate that our algorithm is superior to others in most datasets from the viewpoint of classification accuracy rate. Due to the characteristic in generating linguistic rules, our algorithm is inferior to others in all datasets in terms of rule number. Otherwise, we can summarize the following conclusions through experiment results: the performance of proposed algorithm is better in datasets including two ethnicities than that including three ones; the datasets belonging to male is more suitable to our algorithm than that belonging to female.

5. Conclusion

In this paper, a semantic description method for face features of larger Chinese ethnic groups based on improved WM method are proposed, which is applied to realize learning ethnicity from face. Firstly, an efficient face landmarks detection method (STASM) is adopted to extract face feature points, and corresponding relation between these points and that used in craniofacial research. Secondly, geometric features including distance, area and perimeter are built based on obtained landmarks. Thirdly, linguistic labels are assigned for every geometric features and improved WM method is applied to generate linguistic rules from raw data. Finally, the simplified linguistic rules are used for ethnicity recognition with facial image, which scores the balance between classification accuracy rate and the interpretability of classification result compared with other classification methods. Although the experiment results of our method are promising, there are still insufficiencies to be improved in future work, such as increasing the precision and landmark number of facial feature point with state-of-the-art detection methods (such as [32–34]) and boosting the stability of proposed algorithm.

Acknowledgments

This work is supported by Natural Science Foundation of China (Nos. 61370146, 61175041) and Liaoning Science & Technology of Liaoning Province of China (Nos. 2013405003, 2013020026).

References

[1] L.G. Farkas, M.J. Katic, C.R. Forrest, International anthropometric study of facial morphology in various ethnic groups/races, J. Craniofac. Surg. 16 (2005) 615–646.
[2] Z. Zhuang, D. Landsittel, S. Benson, R. Roberge, R. Shaffer, Facial anthropometric differences among gender, ethnicity, and age groups, Ann. Occup. Hyg. 54 (2010) 391–402.
[3] F. Fang, P. Clapham, K. Chung, A systematic review of interethnic variability in facial dimensions, Plast. Reconstr. Surg. 127 (2011) 874–881.
[4] S. Ritz-Timme, P. Gabriel, J. Tutkuviene, P. Poppa, Z. Obertová, D. Gibelli, D. De Angelis, M. Ratnayake, R. Rizgeliene, A. Barkus, et al., Metric and morphological assessment of facial features: a study on three European populations, Forensic Sci. Int. 207 (2011) 239-e1.
[5] L.-X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Trans. Syst. Man Cybern. 22 (1992) 1414–1427.
[6] L.-X. Wang, The wm method completed: a flexible fuzzy system approach to data mining, IEEE Trans. Fuzzy Syst. 11 (2003) 768–782.
[7] Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition, vol. 10, World Scientific, River Edge, USA, 1996.
[8] L.-X. Wang, C. Wei, Approximation accuracy of some neuro-fuzzy approaches, IEEE Trans. Fuzzy Syst. 8 (2000) 470–478.
[9] X. Yang, J. Yuan, J. Yuan, H. Mao, An improved wm method based on pso for electric load forecasting, Expert Syst. Appl. 37 (2010) 8036–8041.
[10] J. Gou, F. Hou, W. Chen, C. Wang, W. Luo, Improving wang–mendel method performance in fuzzy rules generation using the fuzzy c-means clustering algorithm, Neurocomputing 151 (2015) 1293–1304.
[11] Y. Wang, D. Wang, T. Chai, Extraction and adaptation of fuzzy rules for friction modeling and control compensation, IEEE Trans. Fuzzy Syst. 19 (2011) 682–693.
[12] M.R. Prusty, T. Jayanthi, J. Chakraborty, H. Seetha, K. Velusamy, Performance analysis of fuzzy rule based classification system for transient identification in nuclear power plant, Ann. Nucl. Energy 76 (2015) 63–74.
[13] V. López, S. del Río, J.M. Benítez, F. Herrera, Cost-sensitive linguistic fuzzy rule based classification systems under the mapreduce framework for imbalanced big data, Fuzzy Sets Syst. 258 (2015) 5–38.
[14] V. López, S. del Rio, J.M. Benitez, F. Herrera, On the use of mapreduce to build linguistic fuzzy rule based classification systems for big data, in: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2014 , IEEE, Beijing, China, pp. 1905–1912.
[15] M. Mufti, A. Khanam, Fuzzy rule based facial expression recognition, in: International Conference on Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, IEEE, Sydney, NSW, Australia, p. 57.
[16] A. Khanum, M. Mufti, M.Y. Javed, M.Z. Shafiq, Fuzzy case-based reasoning for facial expression recognition, Fuzzy Sets Syst. 160 (2009) 231–250.

[17] P. Moallem, B.S. Mousavi, S.A. Monadjemi, A novel fuzzy rule base system for pose independent faces detection, Appl. Soft Comput. 11 (2011) 1801–1810.
[18] O. Çeliktutan, S. Ulukaya, B. Sankur, A comparative study of face landmarking techniques, EURASIP J. Image Video Process. 2013 (2013) 13.
[19] S. Milborrow, F. Nicolls, Active shape models with SIFT descriptors and MARS, in: VISAPP, 2014.
[20] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Comput. Vis. Image Underst. 61 (1995) 38–59.
[21] S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model, in: Computer Vision–ECCV 2008, Springer, Marseille, France, 2008, pp. 504–513.
[22] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules, Springer, Berlin, Heidelberg, Germany, 2005.
[23] Wikipedia, List of Ethnic Groups in China, 2014.
[24] J. Shi, A. Samal, D. Marx, How effective are landmarks and their geometry for face recognition? Comput. Vis. Image Underst. 102 (2006) 117–133.
[25] O. Ibáñez, L. Ballerini, O. Cordón, S. Damas, J. Santamaría, An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification, Inf. Sci. 179 (2009) 3998–4028.
[26] O. Ibánez, O. Cordón, S. Damas, J. Santamaria, Modeling the skull–face overlay uncertainty using fuzzy sets, IEEE Trans. Fuzzy Syst. 19 (2011) 946–959.
[27] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226–1238.
[28] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The cas-peal large-scale Chinese face database and baseline evaluations, IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. 38 (2008) 149–161.
[29] D. Xiao-dong, W. Cun-rui, L. Xiang-dong, L. Zhi-Jie, W. Jun, Z. Hai-long, Ethnic features extraction and recognition of human faces, in: 2010 2nd International Conference on Advanced Computer Control (ICACC), vol. 2, IEEE, Shenyang, China, pp. 125–130.
[30] L. Zedong, D. Xiaodong, Z. Qingling, A novel survey based on multiethnic facial semantic web, TELKOMNIKA Indones. J. Electr. Eng. 11 (2013) 5076–5083.
[31] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, ACM SIGKDD Explor. Newslett. 11 (2009) 10–18.
[32] A. Liang, W. Liu, L. Li, M. R. Farid, V. Le, Accurate facial landmarks detection for frontal faces with extended tree-structured models, in: 2014 22nd International Conference on Pattern Recognition (ICPR), IEEE, Stockholm, Sweden, pp. 538–543.
[33] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, pp. 1867–1874.
[34] S. Ren, X. Cao, Y. Wei, J. Sun, Face alignment at 3000 fps via regressing local binary features, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, pp. 1685–1692.

Yuangang Wang received the B.S. and the M.S. degrees in Software Engineering and Computer Technology from Dalian Nationalities University and Dalian Maritime University, Dalian, PR China, in 2012 and 2014, respectively. Currently, he is a Ph.D. candidate in the Research Center of Information and Control, Dalian University of Technology. His research interests include machine learning and data mining.

Xiaodong Duan is the director of Dalian Key Lab of Digital Technology for National Culture, Dalian Nationalities University, Dalian, PR China. He received the B.S. degrees in Computer Science and Technology from Nankai University, Tianjin, PR China in 1985, and received the M. S. and Ph.D. degrees in Applied Mathematics and Computer Software and Theory from Northeastern University, Shenyang, PR China in 1988 and 2001, respectively. His research interests include pattern recognition and data mining.

Xiaodong Liu received the B.S. and the M.S. degrees in Mathematics from Northeastern Normal University in 1986 and Jilin University, Jilin, in 1989, PR China respectively, and the Ph.D. degree in Control Theory and Control Engineering from Northeastern University, Shenyang, PR China in 2003. He is currently a Professor in Research Center of Information and Control, Dalian University of Technology and Department of Applied Mathematics, Dalian Maritime University, a Guest Professor of the ARC Research Center of Excellence in PIMCE, Curtin University of Technology, Australia. He was a Senior Visiting Scientist in Department of Electrical and Computer Engineering, University of Alberta, Edmonton Canada in 2003 and Visiting Research Fellow in Department of Computing, Curtin University of Technology, Perth Australia in 2004. He has been a Reviewer of American Mathematical Reviewer since 1993. He has proposed the AFS theory and is a coauthor of three books. His research interests include algebra rings, combinatorics, topology molecular lattices, AFS (axiomatic fuzzy sets) theory and its applications, knowledge discovery and representations, data mining, pattern recognition and hitch diagnoses, analysis and design of intelligent control systems. Dr. Liu is a recipient of the 2002 Wufu-Zhenhua Best Teacher Award of the Ministry of Communications of People's Republic of China.

Cunrui Wang received the B.S. degrees in Computer Science and Technology from Dalian Nationalities University, Dalian, PR China, in 2002, and received the M.S. degrees in Computer Software and Theory from Northeastern University, Shenyang, PR China, in 2005. Currently, he is a Ph.D. candidate in the Institute of System Science, Northeastern University. His research interests include biometrics identification and data mining.

Zedong Li received the B.S. degree in Automation from Chongqing University of Technology, Chongqing, PR China, in 2006 and received the M.S. degree in Computer Software and Theory from Shenyang Ligong University, Shenyang, PR China, in 2011. Currently, he is a Ph.D. candidate in the Institute of System Science, Northeastern University. His research interests include data mining and fuzzy control.