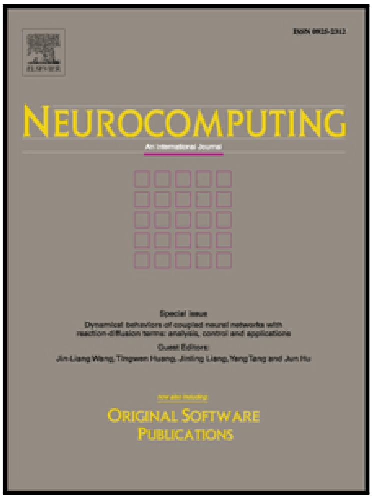
Communicated by Prof. Zidong Wang

Accepted Manuscript



Multi-ethnic Facial Features Extraction based on Axiomatic Fuzzy Set Theory

Zedong Li, Xiaodong Duan, Qingling Zhang, Cunrui Wang, Yuangang Wang, Wanquan Liu

PII: S0925-2312(17)30394-6
DOI: 10.1016/j.neucom.2017.02.070
Reference: NEUCOM 18144

To appear in: *Neurocomputing*

Received date: 8 April 2016
Revised date: 6 February 2017
Accepted date: 25 February 2017

Please cite this article as: Zedong Li, Xiaodong Duan, Qingling Zhang, Cunrui Wang, Yuangang Wang, Wanquan Liu, Multi-ethnic Facial Features Extraction based on Axiomatic Fuzzy Set Theory, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.02.070

Multi-ethnic Facial Features Extraction based on Axiomatic Fuzzy Set Theory

Zedong Li[a,b], Xiaodong Duan[b,*], Qingling Zhang[a], Cunrui Wang[b,a], Yuangang Wang[b,c], Wanquan Liu[d]

[a] *Institute of Systems Science, Northeastern University, Shenyang, Liaoning, China*
[b] *Dalian Key Lab of Digital Technology for National Culture, Dalian Minzu University, Dalian, Liaoning, China*
[c] *Research Center of Information and Control, Dalian University of Technology, Dalian, Liaoning, China*
[d] *Department of Computing, Curtin University, Perth, WA, Australia*

Abstract

This paper proposes a new semantic concept extraction method to choose the salient features for representing multi-ethnic face characteristics based on axiomatic fuzzy set (AFS) theory. It has two advantages, one is that it could well convert the facial features to semantic concepts by bridging the semantic gap between image features and interpretable concepts; the other is that it could be considered as a dimension reduction method to preserve salient features for describing ethnic groups. Firstly, We build facial features to describe face with the landmarks of facial components, such as eyes, mouth and face contour, etc. , and then transform these facial features into semantic concepts. Secondly, a new approach is proposed to obtain the complex semantic concept sets of each ethnic group through clustering simple semantic concept based on AFS framework, and construct an optimal criterion to obtain valid semantic concepts of each ethnic group. Thirdly, we select the typical facial features which are corresponding to the semantic concepts to represent the ethnical face characteristic. Finally, we conduct experiments on Chinese Ethnic Face Database (CEFD), FEI and CK+ database to verify the effectiveness of our method. The K-means and fuzzy c-means(FCM) are used to verify the performance for describing multi-ethnic facial characteristics with the salient facial features. Specially, the obtained results demonstrate the efficacy of our approach, as the semantic concepts generated by optimal model can have an excellent interpretability and comprehension for the facial features. In addition, there is a comparative analysis between our method and other feature selection methods.

Keywords: Semantic Concept Extraction, Axiomatic Fuzzy Set, Facial Features, Ethnical Characteristic

1. Introduction

Facial feature extraction has become one of the most active research areas due to the wide range of applications such as face recognition, face retrieval and ethnic classification. Most traditional approaches, which focus on whole face image, have been developed in the past decades years. Generally, face feature description is divided into two categories, including holistic feature description and local feature description. Holistic feature obtains all information (even all pixels) in face image in each eigenvector, which usually represents facial holistic attribute, such as principal component analysis (PCA) [1, 2, 3], linear discriminant analysis (LDA) [4], independent component analysis (ICA) [5] and discrete cosine transform [6, 7] and other methods [8, 9, 10, 11]. These methods ignore local features. On the contrary, some local feature descriptors have gain significant achievements such as illumination [12], occlusion [13], gender and race recognition [14, 15, 16]. However, they are not able to consider the effect of facial components, which can clearly represent the appearance characteristics including eye, nose, mouth and face contour [17, 18, 19, 20]. Further, the methods of facial attributes [21, 22, 23] pay more attention to facial components, which provides a set of high-level semantic description according to facial attribute such as gender, expression, pose, skin color and so on. However, it is hard to distinguish the similarity of facial components. For instance, the shape of eyebrows are various such as in arc, length, width, which are complicated and difficult to be interpreted. Comparatively, the geometrical

*Corresponding author
Email address: `minzudigital@dlnu.edu.cn` (Xiaodong Duan)

structure could intuitively exhibition their characteristics [24], so that they have been used to match face with the facial geometrical shape including eyebrow arch, zygomatic breadth, chin shape and other geometrical features. Furthermore, geometrical structure method would still generate two disadvantages:(1)the work is hard because the feature points are marked by manual; (2)it is complex to calculate the relationships of geometrical features, which extremely limits the its application. Though, automatic landmark model [25, 26, 27] is proposed, which is widely applied in detecting facial component characteristics, they are usually considered as a set of eigenvector, which could be fused with other holistic or local features for recognizing face [28, 29, 30]. The methods of above mentioned ignore the semantic concepts.

How to valuably and reasonably convert the facial features to semantic concepts is meaningful for representing facial characteristics. Since, the landmark method [31] first is applied in face retrieval, which bridges semantic gap between low-level visual features of image and high-level semantic concepts. In that method, facial components were manually marked, such as the center of the pupils, tip of the nose and the center of the mouth, and then they were transformed to semantic concepts to retrieve a face. Meanwhile, a new method [32] was proposed based on axiomatic fuzzy set (AFS) theory. It establishes semantic concepts, which is more objective and acceptable. The work accomplishes face retrieval with semantic concepts, but it ignores face ethnic attribute. Currently, there are few achievements focusing on Chinese ethnic groups [33]. Therefore, the aim of this paper is to develop a new approach to extract salient features for ethnical groups facial characteristics using the framework of axiomatic fuzzy set (AFS) theory.

Axiomatic fuzzy set (AFS) [34, 35, 36] can objectively convert the facial features to semantic concepts. In the AFS theory, fuzzy sets (membership functions) and their logic operations are algorithmically determined according to the distributions of original data and the semantics of the fuzzy sets. The AFS framework facilitates the studies on how to convert the information in databases into the membership functions and their fuzzy logic operations, by taking both fuzziness (subjective imprecision) and randomness (subjective uncertainty) into account. The advantage of AFS framework is that it does not require to define membership function and initial value as all these are learned from original database. This paper proposes a new semantic concept method to extract the facial features for expressing the characteristics of Chinese ethnic groups such as Mongolian, Korean and Hui. The contributions of our work are summarized follows:

- Using AFS framework, we propose a data-driven algorithm to extract features for representing Chinese multi-ethnic facial characteristics.
- Due to the gender discrepancy, we analyze the facial features corresponding to different genders, and summarize the similarity and differences of facial features respectively.
- We extract the facial features based on the ethnic label. In this study, the aim is to explore whether the facial characteristic is similarity in same ethnic group, when the gender is ignored.

Moreover, the performance of facial features selected by our method is verified using K-means [37, 38] and fuzzy c-means(FCM) [39, 40]. In order to demonstrate the effectiveness of the proposed method, we also conduct some experiments on database of FEI [41, 42] and CK+ [43, 44]. The remainder of this paper is organized as follows: in Sections 2, we give a brief introduction of AFS theory. Section 3 describes how to construct the features with the landmark of facial components. Section 4 details the process of the proposed method, how to select the salient facial features and describe facial characteristics among multi-ethnic groups. Experimental results are reported in Section 5. Finally, we conclude the paper in Section 6.

2. Preliminaries

2.1. AFS algebras

In this section, a segmental facial feature data set(see table 1) of CEFD from literature [33] is used an illustrative example to describe AFS theory. It consists of 10 persons observations and 3 features including Eye Width (f_1), Mouth Width (f_2), and Nose Length(f_3). The facial feature data has been normalized. Let $X = \{x_1, \ldots, x_{10}\}$ be a set of 10 persons observations, where $x_i \in R^3 (i = 1, 2, \ldots, 10)$ denotes the i-th sample, $f_j (j = 1, 2, 3)$ denotes the j-th

Table 1: A segmental facial feature data set with 10 samples

Feature	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Nose length	0.1574	0.1258	0.1953	0.1675	0.4282	0.3316	0.2739	0.4571	0.5005	0.4868
Mouth width	0.1336	0.1462	0.1443	0.1648	0.2716	0.3115	0.2630	0.4791	0.4913	0.4455
Eye width	0.4664	0.4082	0.4751	0.4578	0.3723	0.3901	0.2074	0.6053	0.5385	0.7263

feature of X. Thus, $X = (x_{i,j})$ is a matrix representing data, $x_{i,j}$ is the j-th feature value of x_i. The following three fuzzy IF - THEN rules [35, 36] describe Class 1 for the classification model.

$R1$: IF $x_{i,1}$ is short nose, and $x_{i,2}$ is wide mouth, and $x_{i,3}$ is wide eye, THEN x_i belongs to Class 1;

$R2$: IF $x_{i,2}$ is wide nose, and $x_{i,2}$ is narrow mouth, and $x_{i,3}$ is normal eye, THEN x_i belongs to Class 1;

$R3$: IF $x_{i,1}$ is short nose, and $x_{i,3}$ is wide eye, THEN x_i belongs to Class 1;

Let $M = \{m_{j,k} 1 \leq j \leq 3, 1 \leq k \leq 3\}$ be the set of fuzzy terms, where $m_{j,1}$, $m_{j,2}$, $m_{j,3}$ are fuzzy terms "large", "medium", "small" associated with the feature f_j respectively, then the linguistic fuzzy rules can be written down in the following form:

$R1$: IF x_i is "$m_{1,3}$ and $m_{2,1}$ and $m_{3,1}$", THEN x_i belongs to Class 1;

$R2$: IF x_i is "$m_{1,1}$ and $m_{2,3}$ and $m_{3,2}$", THEN x_i belongs to Class 1;

$R3$: IF x_i is "$m_{1,3}$ and $m_{3,1}$", THEN x_i belongs to Class 1.

For each set of fuzzy terms, $A \subseteq M$; $\prod_{m \in A} m$ represents a conjunction of the fuzzy terms in A. For instance, let $A = \{m_{1,3}, m_{2,1}, m_{1,3}\} \subseteq M$, a new fuzzy set "$m_{1,3}$ and $m_{2,1}$ and $m_{3,1}$" with the linguist interpretation "small nose and large mouth and large eye" can be represented as $\prod_{m \in A} m = m_{1,3}m_{2,1}m_{3,1}$. Then the fuzzy rules can be represented as follows:

$R1$: IF x_i is "$m_{1,3}m_{2,1}m_{3,1}$", THEN x_i belongs to Class 1;

$R2$: IF x_i is "$m_{1,1}m_{2,3}m_{3,2}$", THEN x_i belongs to Class 1;

$R3$: IF x_i is "$m_{1,3}m_{3,1}$", THEN x_i belongs to Class 1.

The antecedent conditions of the three fuzzy rules $R1$, $R2$, $R3$ for cluster 1 can be combined using logic operators "or" as follows:

R: IF x_i is $m_{1,3}m_{2,1}m_{3,1}$ or $m_{1,1}m_{2,3}m_{3,2}$ or $m_{1,3}m_{3,1}$, THEN x_i belongs to cluster 1.

In this case, each rule in R is called an item of R. Here, R has three items (the fuzzy rules $R1$, $R2$, $R3$), so the number of rules is 3. $\sum_{u=1}^{r}(\prod_{m \in A_u} m)$ is a formal sum of the sets $\prod_{m \in A_u} m(A_u \subseteq M)$ is the disjunction of the conjunctions represented by $\prod_{m \in A_u} m$, $u = 1, \ldots, r$. For example, let $A_1 = \{m_{1,3}, m_{2,1}, m_{3,1}\}, A_2 = \{m_{1,1}, m_{2,3}, m_{3,2}\}, A_3 = \{m_{1,3}, m_{3,1}\} \subseteq M$, then a new fuzzy set as the disjunction of $\prod_{m \in A_1} m$, $\prod_{m \in A_2} m$, $\prod_{m \in A_3} m$, i.e., "$m_{1,3}m_{2,1}m_{3,1}$ or $m_{1,1}m_{2,3}m_{3,2}$ or $m_{1,3}m_{3,1}$ can be represented as $\sum_{u=1}^{3}(\prod_{m \in A_u} m) = \prod_{m \in A_1} + \prod_{m \in A_2} + \prod_{m \in A_3}$. Thus, R can be denoted as follows: IF x_i is $\sum_{u=1}^{3}(\prod_{m \in A_u} m)$. THEN x_i belongs to class 1.

The above expressions in R can be formulated as an algebra systems as follows: let M be a non-empty set. The set EM^* is defined as follow:

$$EM^* = \left\{ \sum_{i \in I} (\prod_{m \in A_i} m) | A_i \subseteq M, i \in I, I \ is \ a \ non-empty \ indexing \ set \right\} \tag{1}$$

A binary relation R on EM^* is defined as follows. For any $\sum_{i \in I}(\prod_{m \in A_i} m)$, $\sum_{j \in J}(\prod_{m \in B_j} m) \in EM^*$, $\left[\sum_{i \in I}(\prod_{m \in A_i} m)\right] R \left[\sum_{j \in J}(\prod_{m \in B_j} m)\right] \Leftrightarrow$ (1) $\forall i \in I, \exists h \in J$, such that $A_i \supseteq B_h$; (2) $\forall j \in J, \exists k \in I$,such that $B_j \supseteq A_k$; It is apparent that R is an equivalence relation. The quotient set, EM^*/R is denoted by EM. The notation $\sum_{i \in I}(\prod_{m \in A_i} m) = \sum_{j \in J}(\prod_{m \in B_j} m) \in EM^*$ means that $\sum_{i \in I}(\prod_{m \in A_i} m) = \sum_{j \in J}(\prod_{m \in B_j} m) \in EM^*$ are equivalent under R. Thus the semantics they represent are equivalent. By a straight forward comparison of $m_{1,1}m_{2,1}m_{3,3}$ and $m_{1,1}m_{3,3}$, we conclude that "$m_{1,3}m_{2,1}m_{3,1} + m_{1,1}m_{2,2}m_{3,2} + m_{1,3}m_{3,1}$" and "$m_{1,3}m_{2,1}m_{3,1} + m_{1,3}m_{3,1}$" are equivalent. For any x, the degree of x belonging to the fuzzy set represented by "$m_{1,3}m_{2,1}m_{3,1}$" is always less than or equal to the degree of x belonging to the fuzzy set represented by $m_{1,3}m_{3,1}$. Therefore, the term $m_{1,1}$, $m_{2,1}$, $m_{3,3}$ is redundant in fuzzy set and the expressions "$m_{1,3}m_{2,1}m_{3,1} + m_{1,1}m_{2,2}m_{3,2} + m_{1,3}m_{3,1}$" and "$m_{1,3}m_{2,1}m_{3,1} + m_{1,3}m_{3,1}$" are equivalent in semantics.

In [36], it is proved that $(EM, \bigvee, \wedge)$ is a completely distributive lattice if the lattice operators "$\vee$" and "$\wedge$" are defined as follows: for any fuzzy sets $\sum_{i\in I}(\prod_{m\in A_i} m), \sum_{j\in J}(\prod_{m\in B_j} m) \in EM^*$.

$$\sum_{i\in I}\Big(\prod_{m\in A_i} m\Big) \bigvee \sum_{j\in J}\Big(\prod_{m\in B_J} m\Big) = \sum_{k\in I\sqcup J}\Big(\prod_{m\in B_j} m\Big) \tag{2}$$

$$\sum_{i\in I}\Big(\prod_{m\in A_i} m\Big) \bigwedge \sum_{j\in J}\Big(\prod_{m\in B_J} m\Big) = \sum_{i\in I, j\in J}\Big(\prod_{m\in B_j} m\Big) \tag{3}$$

where for any $k \in I \sqcup J$ (the disjoint union of I and J, i.e., every element in I and every element in J are always regarded as different elements in $I \sqcup J$), $C_k = A_k$ if $k \in I$, and $C_k = B_k$ if $k \in J$.

2.2. Coherence membership functions of fuzzy sets

Let X be a data set and M be a set of fuzzy terms on X. For $A \subseteq M$, $x \in X$, we define

$$A^{\succeq}(x) = \{y \in X | x \succeq_m y \ for\ any\ m \in A\} \subseteq X \tag{4}$$

where a linearly ordered relation is denoted by "$\succeq$". For $m \in M$, "$x \succeq_m y$" implies that the degree of x belonging to m is larger than or equal to that of y. $A^{\succeq}(x)$ is the set of all elements in whose degrees of belongingness to set $\prod_{m\in A} m$ are less than or equal to that of x. $A^{\succeq}(x)$ is determined by the semantics of fuzzy set A and the probability distribution of observed data set X.

For fuzzy set $\xi \in EM$, let $\mu_\xi : X \to [0, 1]$. $\{\mu_\xi(x) | \xi \in EM\}$ is called a set of coherence membership functions of the AFS fuzzy logic system $(EM, \wedge, \vee)$[27], if the following conditions are satisfied.

1. For $\alpha, \beta \in EM$, if $\alpha \le \beta$ in lattice $(EM, \wedge, \vee)$, then $\mu_\alpha(x) \le \mu_\beta(x)$ for any $x \in X$;
2. For $x \in X, \eta = \sum_{i\in I}()$, if $A_i^{\succeq}(x) = \phi$ for all $i \in I$ then $\mu_\eta(x) = 0$;
3. For $x, y \in X, A \subseteq M, \eta = \sum_{i\in I}(\prod_{m\in A_i}) \in EM$, if $A^{\succeq}(x) \subseteq A^{\succeq}(y)$, then $\mu_\eta(x) \le \mu_\eta(y)$; if$A^{\succeq}(x) = X$ then $\mu_\eta(x) = 1$.

The coherence membership functions are associated with a measure over X. We propose two types of measures for fuzzy sets, which can be constructed by taking the semantics of the fuzzy terms and the probability distribution of the feature of the data into account. In order to achieve this, we first introduce the following definition.

Definition 1 ([35]). *Let v be a fuzzy term on X. $\rho_v : X \to R^+ = [0, \infty)$. ρ_v is called a weight function of the simple concept v if ρ_v satisfies the following conditions:*

1. *For $x \in X$, $\rho_v(x) = 0$, if x does not belong to v;*
2. *For $x, y \in X$, $\rho_v(x) \ge \rho_v(y)$, if the degree of x belonging to v is larger than or equal that of y.*

In what follows, we present how to define the coherence membership functions in a probability measure space.

Theorem 1 ([36]). *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability measure space and M be a set of fuzzy terms on X. Let ρ_γ be the weight function for a fuzzy term $\gamma \in M$. Let $X \subseteq \Omega$ be a finite set of observed samples from the probability space $(\Omega, \mathcal{F}, \mathcal{P})$. If for any $m \in M$ and any $x \in \Omega$; $m^{\succeq}(x) \in \mathcal{F}_m$. Then the following assertions hold:*

1. *$\{\mu_\xi(x) | \xi \in EM\}$ is a set of coherence membership functions of $(EM, \wedge, \vee)$, provided that the membership function for each fuzzy set $\xi = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$ is defined as follows:*

$$\mu_\xi(x) = \sup_{i\in I} \inf_{\gamma\in A_i} \frac{\sum_{u\in A_i^{\succeq}(x)} \rho_\gamma(u) N_u}{\sum_{u\in X} \rho_\gamma(u) N_u}, \forall x \in X \tag{5}$$

$$\mu_\xi(x) = \sup_{i\in I} \inf_{\gamma\in A_i} \frac{\int_{u\in A_i^{\succeq}(x)} \rho_\gamma(u) N_u}{\int_{u\in X} \rho_\gamma(u) N_u}, \forall x \in \Omega \tag{6}$$

2. *If for every $\gamma \in M, \rho_\gamma(x)$ is continuous on Ω and X is a set of samples randomly drawn from the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, then the membership function defined by (5) converges to the membership function defined by (6), for all $x \in X$ as $|X|$ approaches infinity.*

Theorem 1 defines the membership functions based on the fuzzy logic operations expressed on the observed data and the overall space by taking both fuzziness and randomness into account via $\rho_\gamma(x)$ and $A_i^{\succ}(x)$. The following practical relevance of the coherence membership functions can be ensured by Theorem 1.

- The membership functions and the fuzzy logic operations determined by the observed data drawn from a probability space will be consistent with those being determined by the probability distribution expressed in the probability space.
- The results obtained via the AFS fuzzy logic based on the membership functions and their logic operations determined by different data sets drawn from the same probability space will be consistent.
- The laws discovered based on the membership functions and their logic operations determined by the observed data drawn from a probability space can be applied to the whole space by using the membership.

The mathematical properties of AFS structure and AFS algebra are discussed in literature [34, 35, 36], and offers a comprehensive introduction of the AFS theory and its applications.

3. Facial characteristic and semantic concept construction

There are mainly two steps for obtaining the facial features. The first step is to detect the landmarks of facial components, such as eyes, eyebrows, nose, based on landmark model [25]. The second step is to construct the facial feature using the landmarks. The last one is to build semantic concepts for describing facial features.

Let $I^{face} = \{I_1, I_2, \ldots, I_h\}$ be a set of frontal face images, and assume that all expression is neutral on face image. The landmark model is used to detect the facial landmarks, which are labeled by numbers as shown Figure 1. Respectively, we extract 76 landmarks sets form each face image $I_k \in I^{face}$. $L_k^{face} = \{l_1, l_2, \ldots, l_{76}\}$ is a 76 landmarks set about k-th face image, $l_i = (x_i, y_i) \in L$ represents the characteristic point of facial component. x_i and y_i are x-axis and y-axis coordinates of i-th point l_i. In this case, we assume that the characteristic points describing the facial components such as eyes, nose, mouth and face contour, are extracted. Let $F_h = \{f_1, f_2, \ldots, f_j\}$ be a facial feature vector I_h built by facial characteristic points. I_h^F represents the facial features on the k-th face image I_h.

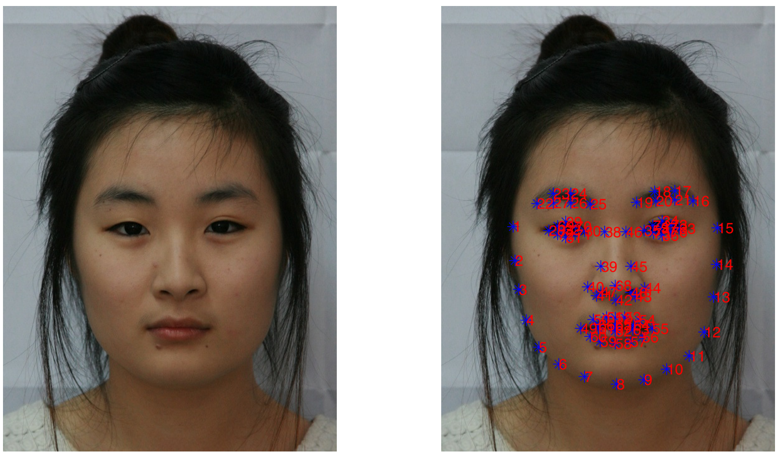


Figure 1: The landmarks of facial components based on landmark model

In [32], it only builds the facial features about eyes, mouth and nose using landmarks. Due to independence of facial components, the facial characteristics do not be well represented. In contrast to it, we increase some new facial features for describing facial characteristics in detail. For example, in Figure 1, the face image illustrated is a Korean female. It marks 76 points of facial components, such as eyes, brows, mouth, nose and facial contour, etc.. There are many facial features for describing facial characteristics using these landmark points. The features are illustrated in Table 2.

Table 2: Facial features illustration

Feature f_j	The calculation of facial feature
f_1	$d(l_{34}, l_{36})$
f_2	$d(l_{33}, l_{35})$
f_3	$\arccos\{[d^2(l_{33}, l_{34}) + d^2(l_{34}, l_{35}) - d^2(l_{33}, l_{35})]/[2 * d(l_{33}, l_{34} * d(l_{34}, l_{35}))]\}$
f_4	$\arccos\{[d^2(l_{34}, l_{35}) + d^2(l_{35}, l_{36}) - d^2(l_{34}, l_{36})]/[2 * d(l_{34}, l_{35} * d(l_{35}, l_{36}))]\}$
f_5	$\arccos\{[d^2(l_{33}, l_{34}) + d^2(l_{33}, l_{36}) - d^2(l_{34}, l_{36})]/[2 * d(l_{33}, l_{34} * d(l_{33}, l_{36}))]\}$
f_6	$d(l_{29}, l_{31})$
f_7	$d(l_{28}, l_{30})$
f_8	$\arccos\{[d^2(l_{28}, l_{29}) + d^2(l_{29}, l_{30}) - d^2(l_{28}, l_{30})]/[2 * d(l_{28}, l_{29} * d(l_{29}, l_{30}))]\}$
f_9	$\arccos\{[d^2(l_{29}, l_{30}) + d^2(l_{30}, l_{31}) - d^2(l_{29}, l_{31})]/[2 * d(l_{29}, l_{30} * d(l_{30}, l_{31}))]\}$
f_{10}	$\arccos\{[d^2(l_{28}, l_{29}) + d^2(l_{28}, l_{31}) - d^2(l_{29}, l_{31})]/[2 * d(l_{28}, l_{29} * d(l_{28}, l_{31}))]\}$
f_{11}	$d(l_{32}, l_{37})$
f_{12}	$d(l_{40}, l_{44})$
f_{13}	$d(c(l_{38}, l_{46}), l_{42})$
f_{14}	$d(l_{68}, l_{42})$
f_{15}	$d(c(l_{38}, l_{46}), l_{68})$
f_{16}	$d(l_{38}, l_{39}) + d(l_{39}, l_{40}) + d(l_{40}, l_{41}) + d(l_{41}, l_{42}) + d(l_{42}, l_{43}) + d(l_{43}, l_{44}) + d(l_{44}, l_{45}) + d(l_{45}, l_{46}) + d(l_{46}, l_{38})$
f_{17}	$area(l_{38}, l_{39}, \ldots, l_{46})$
f_{18}	$d(l_1, l_{15}) + d(l_2, l_{14}) + d(l_3, l_{13})$
f_{19}	$d(c(l_{38}, l_{46}), l_8)$
f_{20}	$d(l_{49}, l_{55})$
f_{21}	$d(l_{52}, l_{58})$
f_{22}	$d(l_{49}, l_{50}) + d(l_{50}, l_{51}) + d(l_{51}, l_{52}) + d(l_{52}, l_{53}) + d(l_{53}, l_{54}) + d(l_{54}, l_{55}) + d(l_{55}, l_{56}) + d(l_{56}, l_{57}) + d(l_{57}, l_{58}) + d(l_{58}, l_{59}) + d(l_{59}, l_{60}) + d(l_{60}, l_{49})$
f_{23}	$area(l_{49}, l_{50}, \ldots), l_{60}$
f_{24}	$d(l_8, l_{58})$
f_{25}	$\arccos\{[d^2(l_4, l_8) + d^2(l_8, l_{12}) - d^2(l_4, l_{12})]/[2 * d(l_4, l_8) * d(l_8, l_{12}))]\}$
f_{26}	$\arccos\{[d^2(c(l_{25}, l_{19}), l_2) + d^2(c(l_{25}, l_{19}), l_{14}) - d^2(l_2, l_{14})]/[2 * d(c(l_{25}, l_{19}), l_2) * d(c(l_{25}, l_{19}), l_{14})]\}$
f_{27}	$\arccos\{[d^2(l_{22}, l_{42}) + d^2(l_{16}, l_{42}) - d^2(l_{22}, l_{16})]/[2 * d(l_{22}, l_{42}) * d(l_{16}, l_{42})]\}$
f_{28}	$\arccos\{[d^2(c(l_{25}, l_{19}), l_{47}) + d^2(c(l_{25}, l_{19}), l_{48}) - d^2(l_{47}, l_{48})]/[2 * d(c(l_{25}, l_{19}), l_{47}) * d(c(l_{25}, l_{19}), l_{48})]\}$
f_{29}	$\arccos\{[d^2(c(l_{25}, l_{19}), l_{32}) + d^2(c(l_{25}, l_{19}), l_{37}) - d^2(l_{32}, l_{37})]/[2 * d(c(l_{25}, l_{19}), l_{32}) * d(c(l_{25}, l_{19}), l_{37})]\}$

In Table 2, $d(l_i, l_j)(i, j \in L_h^{face})$ denotes the Euclidean distance between point l_i and l_j; The angle feature Arc is calculated by arc-cosine function; $c(l_i, l_j)$ presents the central point between l_i and l_j. Their formula is respectively as follows

$$\begin{cases} d(l_i, l_j) = \left\|l_i - l_j\right\|_2 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\ Arc = \arccos\left[d^2(l_i, l_j) + d^2(l_s, l_r) - d^2(l_p, l_q)\right]/[2 * d(l_i, l_j) * d(l_s, l_r)] \\ c(l_i, l_j) = (\frac{1}{2}(x_i + x_j), \frac{1}{2}(y_i + y_j)) \end{cases}$$

$area(l_i, \ldots, l_j)$ shows the area which surrounds the facial region from l_i to l_j in order. So, I_k^f represents the facial feature $f \in F$ on k-th face image I_k. For example, when f_1= "eyes height", it is eyes height on k-th face image I_h. In addition, due to the facial symmetry, the feature f_1 and f_2 contains right eye and left eye. Then, the facial features F_k on face image I_h are defined, we give semantic concept terms $m_{j,k}$ to represent facial features $f_j \in F$, which are shown in Table 3.

As an example of facial feature f_1("eyes height") in Table 3, $m_{1,1}, m_{1,2}, m_{1,3}$ are semantic concepts term, "Large",

Table 3: Semantic concept term $m_{j,k}$ for each feature f_j

Feature f_j	Introduction of facial feature	Semantic concept term, $m_{j,k}$
f_1	Left eye height	
f_2	Left eye width	
f_3	The angle between left eyes canthus and left eyes top point	
f_4	The angle between left eyes top and down point and left eyes inner canthus	
f_5	The angle between left eyes top and down point and left eyes outer canthus	
f_6	Right eye height	
f_7	Right eye width	
f_8	The angle between right canthus and right eye top point	
f_9	The angle between right eyes top and down point and right eyes inner canthus	
f_{10}	The angle between right eyes top and down point and right eyes outer canthus	
f_{11}	Pupils distance	
f_{12}	Nose width	
f_{13}	Nose length	
f_{14}	Nose tip height	
f_{15}	Nasal peak height	Large,Medium,Small
f_{16}	Nose perimeter	
f_{17}	Nose area	
f_{18}	Face width	
f_{19}	Face length	
f_{20}	Mouth width	
f_{21}	Mouth height	
f_{22}	Mouth perimeter	
f_{23}	Mouth area	
f_{24}	Underjaw height	
f_{25}	The angle between zygomatic and underjaw apex	
f_{26}	The angle between zygomatic and eyebrow centre	
f_{27}	The angle between outer eyebrows and nose apex	
f_{28}	The angle between nostrils and eyebrow centre	
f_{29}	The angle between eyebrows centre and pupils	

"Medium", and "Small" associated with the facial feature f_1. The semantic concepts of "Large", "Medium" and "Small" are obtained by setting threshold value.

4. Facial characteristic description of ethnic groups via semantic concepts

From [33], we realize that the face may have multi-characteristics. That is that exist a few facial features, which can reveal the differences of multi-ethnics. Therefore, we analyze the correlation among facial features and searching reasonable semantic concepts to interpret the facial characteristic for ethnic groups. In this section, we propose a new method of ethnic facial characteristic description based on AFS theory to select effective facial features, and reveal differences of multi-ethnic face.

There are three steps detailed to select the best optimal semantic concepts for each ethnic group. The first step is to extract valid simple semantic concepts, which represent the facial characteristic in Table 2. The second step is to build reasonable combined semantic concepts from F to describe each ethnic group. The last step is to search the best optimized the semantic concept sets for facial characteristic of each ethnic group. The key symbols are listed in Table 4.

Table 4: Key symbols in this paper

Symbol	Description of symbol
X	A set of all facial features training samples
X_{C_i}	A set of i-th class facial features training samples in X
C_i	The i-th class facial features training samples in X
N	The number of all facial features training samples in X
N_C	The number of class in X
N_{C_i}	The number of facial features training samples X_{C_i}
F	A facial features sets, which describes the facial characteristic
f_j	The j-th facial feature in F
$m_{j,k}$	The k-th semantic concept associated with the j-th facial feature
M_j	A set of semantic concepts for j-th facial feature
$M_j^{x_i}$	A semantic sets selected for describing individual x_i in X
EM_{x_i}	A conjunction semantic sets selected for describing invididual x_i in X
ξ_{c_i}	A semantic concept sets for C_i class
δ_1	A threshold value for selecting semantic concept sets of $M_j^{x_i}$
δ_2	A threshold value for selecting conjunction semantic EM^{x_i}

Let X be a set of facial feature training samples. According to Table 3, we know $F = \{f_1, f_2, \ldots, f_{17}\}$ is a set of facial features including 29 facial characteristics, $M_j = \{m_{j,k} 1 \leq j \leq 29, 1 \leq k \leq 3\}$ is a set of semantic concept term, where $m_{j,k}$ is semantic concept associated with the feature f_j in F. The membership function, determined by the given facial characteristics database, of any semantic concept $m_{j,k} \in M_j$ is defined by formula (5). In follows, we will detail the process.

4.1. Select the salient semantic concept sets for each individual

After defining the semantic concept, we define the membership function of facial feature f_j belonging to semantic term $m_{j,k} \in M_j$. The membership function is determined by the given database, and it represents each image with some semantic terms. For individual $x_i \in X_{C_i}$, we select available simple semantic concept $m_{j,k}$ for x_i.

$$M_j^{x_i} = \{m_{j,k} | \mu_{m_{j,k}}(x_i) \geq \max(\mu_{m_{j,k}}(x_i)) - \delta_1\} \tag{7}$$

$$\mu_{j,k}(x_i) = \sup_{i \in X} \inf_{m \in M} \frac{\sum_{u \in m_{j,k}} \rho_m(u) N_u}{\sum_{u \in X} \rho_m(u) N_u}, \forall x_i \in X_{C_i}$$

In which, $\mu_{m_{j,k}}(x_i)$ represents the membership degree of x_i belonging to $m_{j,k}$. δ_1 is a threshold value for recognizing semantic concepts. We seek the maximum $\mu_{m_{j,k}}(x_i)$ of all $m_{j,k}$. Then, for each individual $x_i \in X_{C_i}$, we can extract a valid semantic set $M_j^{x_i}$.

Then, we can describe individual x_i using a complex semantic concept A_i, which is the conjunctions of the $m_{j,k}$ in of M^{x_i}. Along with the A_i generated, there are two disadvantages: One is computational complexity of semantic concept will increase. For a semantic concept $\xi = \prod_{m\in H} m$, if the number of the semantic concept terms m is larger, the description for objects is more particular. But, the computational complexity of objects will become complicated. In order to solve it, we set a threshold value δ_2 to judge whether the semantic concept $A_i = \prod_{m\in H} m$ is remained. If the membership degree of A_i belonging to objects is greater than or equal to δ_2, we will reserve A_i, which describes the objects well. On the contrary, we will abandon semantic concept A_i.

$$EM^{x_i} = \left\{ \sum A_i | A_i = \prod_{m_{j,k}\in H} m_{j,k}, H \subseteq M_{x_i}, \mu_{A_i}(x_i) \geq \delta_2 \right\}$$

Where, $\mu_{A_i}(x_i)$ is membership degree of x_i belonging to A_i. δ_2 is another threshold value for obtaining conjunction semantic concepts of x_i.

Another difficultly is that it will generate a mass of redundant semantic concepts. Along with the semantic concept A_i being larger, the membership degrees of objects belonging to objects may be too small (may be close to 0). The semantic concept $A_i = \prod_{m\in H} m$ has weak expression for objects so that it became redundancy. So, we set an evaluation index $V_{A_i}^{x_i}$ to seek the best semantic concept A_i in describing individual x_i.

$$\max V_{A_i}^{x_i} = E_{A_i}^{dis}(x_i) \cdot T_{A_i}^{dis}(x_i) \cdot W_{A_i}^{dis}(x_i) \tag{8}$$

$$s.t. \begin{cases} E_{A_i}^{dis}(x_i) = \dfrac{\overset{(x_i)}{E_{A_i}^{dis}}}{\sum_{A_i\in EM^{x_i}} \overset{(x_i)}{E_{A_i}^{dis}}}, \overset{(x_i)}{E_{A_i}^{dis}} > 0 \\ T_{A_i}^{dis}(x_i) = \dfrac{\overset{(x_i)}{T_{A_i}^{dis}}}{\sum_{A_i\in EM^{x_i}} \overset{(x_i)}{T_{A_i}^{dis}}}, \overset{(x_i)}{T_{A_i}^{dis}} > 0 \\ 1 \geq W_{A_i}^{dis}(x_i) \geq 0 \end{cases}$$

where, $V_{A_i}^{x_i}$ is a scalar value for choosing semantic concepts belonging to x_i. This evaluation function consists of three parts, $E_{A_i}^{dis}(x_i)$, $T_{A_i}^{dis}(x_i)$ and $W_{A_i}^{dis}(x_i)$. Each having a clear interpretation as follow.

$$\begin{aligned} \overset{(x_i)}{E_{A_i}^{x_i}} &= \frac{\sum\limits_{C=C_1, C\neq C_i}^{N_C} \sum\limits_{x=x_1, x\notin X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2}{\sum\limits_{x=x_1, x\in X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2} \\ &= \frac{\sum\limits_{C=C_1, C\notin C_i}^{N_C} \sum\limits_{x=x_1, x\notin X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2 + \sum\limits_{x=x_1, x\in X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2 - \sum\limits_{x=x_1, x\in X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2}{\sum\limits_{x=x_1, x\in X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2} \\ &= \frac{\sum\limits_{C=C_1}^{N_C} \sum\limits_{x=x_1, x\notin X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2}{\sum\limits_{x=x_1, x\in X_{C_i}}^{N_{C_i}} \left\| \mu_{A_i}(x_i) - \overline{\mu_{A_i}}(X_{C_i}) \right\|^2} - 1 \end{aligned} \tag{9}$$

$$\mu_{A_i}(x_i) = \sup_{i\in X} \inf_{m\in A_i} \frac{\sum_{u\in A^{\succ}(x_i)} \rho_m(u) N_u}{\sum_{u\in X_{C_i}} \rho_m(u) N_u}, \forall x_i \in X_{C_i}$$

$$\overline{\mu_{A_i}}(X_{C_i}) = \frac{1}{N_{C_i}} \sum_{x=x_1}^{N_{C_i}} \mu_{A_i}(x_i) = \frac{1}{N_{C_i}} \sum_{x=x_1}^{N_{C_i}} \sup_{i\in X} \inf_{m\in A_i} \frac{\sum_{u\in A^{\succ}} \rho_m(u) N_u}{\sum_{u\in X_{C_i}} \rho_m(u) N_u}, \forall x_i \in X_{C_i}$$

where, $E^{dis}_{A_i}(x_i)$is a measure function, which indicates the dispersion among different class X_{C_i}. It requires within-class distance is minimized while the between-class distance is maximized with the semantic concept A_i. Thus, we want to achieve the balance of these objectives by maximizing $E^{dis}_{A_i}$. The larger $E^{dis}_{A_i}(x_i)$, the more important A_i for the semantic concept of individual x_i i.e.. $\mu_{A_i}(x_i)$ is the membership degree of x_i belonging to A_i. $\overline{\mu_{A_i}}(X_{C_i})$ is the mean membership degree of $x_i \in X_{C_i}$ belonging to $A_i \in EM^{x_i}$.

$$T^{dis}_{A_i}(x_i) = 1 - \frac{\sum_{x_i\in X_{C_i}} |\mu_{x_i}(A_i) - \overline{\mu_{X_{C_i}}}(A_i)|}{N_{C_i}} \tag{10}$$

$$\mu_{x_i}(A_i) = \sup_{i\in X} \inf_{A_i\in EM^{x_i}} \frac{\sum_{u\in A^{\succ}(x_i)} \rho_m(u) N_u}{\sum_{u\in X_{C_i}} \rho_m(u) N_u}, \forall x_i \in X_{C_i}, A_i \in EM^{x_i}$$

$$\overline{\mu_{X_{C_i}}}(A_i) = \frac{1}{N_{C_i}} \sum_{x=x_1}^{N_{C_i}} \mu_{X_{C_i}}(A_i) = \frac{1}{N_{C_i}} \sum_{x=x_1}^{N_{C_i}} \sup_{i\in X} \inf_{A_i\in EM^{x_i}} \frac{\sum_{u\in A^{\succ}} \rho_m(u) N_u}{\sum_{u\in X_{C_i}} \rho_m(u) N_u}, \forall x_i \in X_{C_i}, A_i \in EM^{x_i}$$

In this value, $T^{dis}_{A_i}$ is another measure function, it represents close degree between A_i and semantic concept sets EM_{x_i}. The larger, the closer A_i for EM^{x_i}. N_{C_i} is the number of samples X_{C_i}. The means of $\mu_{x_i}(A_i)$ and $\overline{\mu_{X_{C_i}}}(A_i)$ is as similar as $\mu_{A_i}(x_i)$ and $\overline{\mu_{A_i}}(X_{C_i})$. Therefore, $T^{dis}_{A_i}(x_i)$ is used to measure whether A_i is close to the set EM^{x_i}, which belongs to x_i.

$$W^{dis}_{A_i} = \frac{\mu_{A_i}(x_i)}{\sum_{A_i\in EM^{x_i}} \mu_{A_i}(x_i)} \tag{11}$$

$$\mu_{A_i}(x_i) = \sup_{i\in X} \inf_{m\in A_i} \frac{\sum_{u\in A^{\succ}(x_i)} \rho_m(u) N_u}{\sum_{u\in X_{C_i}} \rho_m(u) N_u}, \forall x_i \in X_{C_i}$$

where, $W^{dis}_{A_i}$ is a weight function for $A_i \in EM^{x_i}$. It represents the measures the weight of semantic concept $A_i \in EM^{x_i}$. The weight of A_i is larger, the better description for A_i, which is useful for us.

According to process referred above, we will combine three evaluation values including $E^{dis}_{A_i}(x_i)$,$T^{dis}_{A_i}(x_i)$ and $W^{dis}_{A_i}(x_i)$ to estimate each A_i for each x_i. Finally, the maximum value of $V^{x_i}_{A_i}$ corresponding to A_i for x_i will be selected.

4.2. Facial characteristic description of each ethnic groups

Now, we have obtained a best semantic concept A_i for individual x_i using $V^{x_i}_{A_i}$. Then, we can construct description for C_i class with the semantic concept A_i for each x_i.

$$EM^{C_i} = \left\{ \sum_{x_i\in X_{C_i}} A^{x_i}_i | V^{x_i}_{A_i}(A_i) = \max\{V^{x_i}_{A_i}\}, A^{x_i}_i \subseteq EM^{x_i} \right\} \tag{12}$$

In this study, we exclude large number of insignificant semantic concepts of x_i. We obtain the facial semantic concept sets of each ethnic group using all x_i belonging to X_{C_i}. The semantic concept sets remained for C_i are more accurate, compact, interpretable and understandable. And the computation complexity will be greatly reduced by the application of some optimization-based approaches. Due to the similarity of face in same ethnic group and differences of that among multi-ethnic groups, we will give some new restraining conditions to ensure the salient features of each ethnic group.

4.3. Optimize the semantic concept set for selecting salient features of each ethnic group

One can see that $V_{A_i}^{x_i}$ is larger, the more valuable the semantic concept for x_i we can have obtain semantic concept sets EM^{C_i} for each class C_i, which contains abundant facial features. However, a best semantic concept of C_i, not only the differences among multi-ethnic groups should be as large as possible, but also the similarity of x_i in X_{C_i} should be as close as possible. The facial features among classes can be as clear as possible. For this purpose, we set three restraining conditions such as $G^{C_i}(\omega_j)$,$Q^{C_i}(\omega_j)$ and $G^{C_i}(\omega_j)$, for the salient facial features of class C_i.

$$\max \xi_{C_i}(\omega_j) = G^{C_i}(\omega_j) \bigvee Q^{C_i}(\omega_j) \bigvee G^{C_i}(\omega_j), \omega_i \in EM^{C_i} \tag{13}$$

$$s.t. \begin{cases} G^{C_i}(\omega_i) = \arg \max\limits_{\omega_i \in EM^{C_i}} \{R_{\omega_i}^{C_i}\} \\ Q^{C_i}(\omega_i) = \arg \max\limits_{\omega_i \in EM^{C_i}} \{S_{\omega_i}^{C_i}\} \\ P^{C_i}(\omega_i) = \arg \max\limits_{\omega_i \in EM^{C_i}} \{D_{\omega_i}^{C_i}\} \end{cases}$$

In formula (13), The symbol "$\vee$" represents the logical relationship "or", it means that the semantic concepts by chosen should conform the restraining condition simultaneously. Finally, we will combine these semantic concepts as a description of the facial characteristics for the C_i-th ethnic group. where, the $R_{\omega_i}^{C_i}$ indicates the sum of membership degree of all $x_i \in X_{C_i}$ belonging to each semantic concept $\omega_i \in EM^{C_i}$. Its presentation is as follows:

$$R_{\omega_i}^{C_i} = \sum_{x=x_1}^{N_{C_i}} \mu_{\omega_i}(x_i) \tag{14}$$

$$\mu_{\omega_i}(x_i) = \sup_{i \in X} \inf_{m \in \omega_i} \frac{\sum_{u \in \omega^{>}(x_i)} \rho_m(u) N_u}{\sum_{u \in X_{C_i}} \rho_m(u) N_u}, \forall x_i \in X_{C_i}, \omega_i \in EM^{C_i} \tag{15}$$

In which, $S_{\omega_i}^{C_i}$ is sum of difference value of membership degree between class C_i and any other class $C_j \in X(C_i \neq C_j)$ on $\omega_i \in EM^{C_i}$. Moreover, we will set a estimation value for each $\omega_i \in EM^{C_i}$. Its detail is described as follows:

$$S_{\omega_i}^{C_i} = \sum_{x_i \in X_{C_i}} \{\mu_{\omega_i}(x_i) - \max\{\mu_{\omega_i}(x_j)\}\}, x_i \in X_{C_i}, x_j \notin X_{C_i}, x_j \in X, \omega_i \in EM^{C_i} \tag{16}$$

where, the computation approach of $\mu_{\omega_i}(x_i)$ and $\mu_{\omega_i}(x_j)$ is as similar as formula (15).

$D_{\omega_i}^{C_i}$ is last measure index to estimate the semantic concept $\omega_i \in EM^{C_i}$. If semantic concept ω_i is the best description for class C_i, it will achieve the highest accuracy in semantic concept sets EM^{C_i} on data X. Thus, we take advantage of the result to extract $\omega_i \in EM^{C_i}$.

Finally, we can obtain the the semantic concept sets ξ_{C_i} of class C_i, which is made up by three restrain conditions $R_{\omega_i}^{C_i}$, $S_{\omega_i}^{C_i}$ and $D_{\omega_i}^{C_i}$.

$$\xi_{C_i} = \Bigg\{ \bigvee_{\alpha = \arg\max_{\omega_i \in EM^{C_i}} \{R_{\omega_i}^{C_i}\}} \alpha \Bigg\} \bigvee \Bigg\{ \bigvee_{\beta = \arg\max_{\omega_i \in EM^{C_i}} \{S_{\omega_i}^{C_i}\}} \beta \Bigg\} \bigvee \Bigg\{ \bigvee_{\eta = \arg\max_{\omega_i \in EM^{C_i}} \{R_{\omega_i}^{C_i}\}} \eta \Bigg\} \tag{17}$$

5. Experimental Results

Empirical tests, observations, analysis and evaluations are presented in this section. Three experiments main consist of Chinese Ethnic Face Database (CEFD) [33] database, FEI [41, 42] and CK+ [43, 44]. CEFD includes 535 individuals, which is gathered by our term. It includes Mongolian, Korean, Hui, who are the mainly groups living in North China. These images are from student volunteer in Dalian Minzu university, whose age is from 18 to 22. The number of Mongolian, Korean, Hui is 175(male 88, female 87), 180(male 90, female 90), 180(male 90, female 90). The FEI database contains 2,800 images from 200 different subjects (100 male, 100 female), which are gathered by

Algorithm 1 Optimizing facial semantic concept set for the salient features

Input: F is a set of facial features; X is a set of facial features; P is parameters to divide semantic concept "large, medium and small".
Output: ξ_{c_i} is a semantic concept set of class C_i.

Compute $\mu_{m_{j,k}}(x_i), x_i \in X(i = 1, 2, \ldots, N), m_{j,k} \in M_j^{x_i}(1 \leq j \leq 29, 1 \leq k \leq 3)$;
for each sample x_i in $X(i = 1, 2, \ldots, N)$ **do**
  if $\mu_{m_{j,k}}(x_i) \geq \max \mu_{m_{j,k}}(x_i) - \delta_1$ **then**
    $M_j^{x_i} = \{\sum m_{j,k}\}$;
    if $\mu_{m_{j,k}}(x_i) \geq \delta_2, A_i = \prod m_{j,k}$ **then**
      $EM^{x_i} = \{\sum A_i\}$
    end if
  end if
end for
for each sample x_i in $X(i = 1, 2, \ldots, N)$ **do**
  Compute $V_{A_i}^{x_i} = E_{A_i}^{dis}(x_i) \cdot T_{A_i}^{dis}(x_i) \cdot W_{A_i}^{dis}(x_i)$;
  $A_i^{x_i} = \arg\max V_{A_i}^{x_i}$;
  $EM^{C_i} = \sum_{x_i \in X_{C_i}} A_i^{x_i}$
end for
for each class $C_i(i = 1, 2, \ldots, N_C)$ **do**
  Compute $R_{\omega_j}^{C_i}, S_{\omega_j}^{C_i}, D_{\omega_j}^{C_i}$;
  $\xi_{C_i} = \arg\max R_{\omega_j}^{C_i} \vee \arg\max S_{\omega_j}^{C_i} \vee \arg\max D_{\omega_j}^{C_i}$;
end for

college students of Brazil. All face is upright frontal position with profile rotation of up to about 180 degrees. The age is mainly between 19 and 40 years old. In our experiments, we use 100 the frontal position and neutral expression images of female to test our method. CK+ is an extended face database for CK database. There are 593 sequences across 123 subjects, and all sequences are from the neutral face to the peak expression. We choose 65 neutral faces (American female) subjects using in our experiments.

In order to verify our method, we only use the natural expression of every individual. According to various size of face images, we will normalize all features on basis of the feature $f_h = d(l_{30}, l_{35})$. We use a 5-fold cross validation for all data sets. Each data set is divided into 5-fold of (approximately) equal size. Each time one fold is left out of the whole data set from training, and this fold is used for testing. All membership functions of semantic concept sets belonging to EM are determined by (5) in Theorem 1. For any $m_{j,k} \in M_j$, $\rho_{m_{j,k}}(x_i)$ described by Gaussian function $\rho_{m_{j,k}}(x_i) = e^{-\frac{(x_i-c)^2}{0.2\sigma^2}}$ represents any x_i belonging to $m_{j,k}$ at some degree, where σ^2 is the variance of facial feature f_j, c is a constant to distinguish semantic "large", "medium" , "small". $N_{x_i} = 1$ denotes x_i is observed for one time.

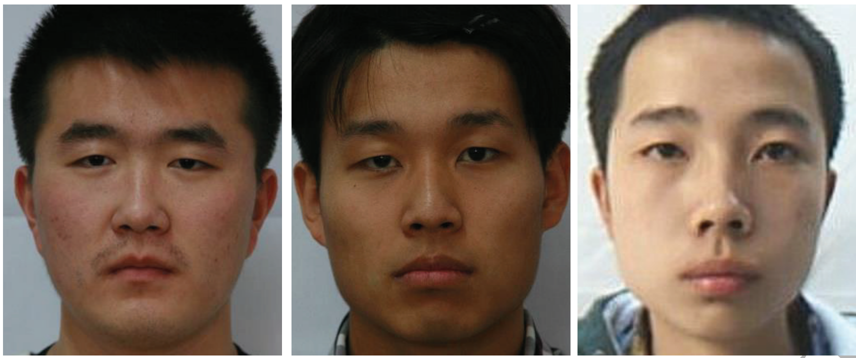
In experiment, three semantic concepts are specialized on each feature f_j. $m_{j,1}$, $m_{j,2}$, $m_{j,3}$ are three semantic concept terms respectively, , "large", "medium", "small" associated with the feature f_j in F. $m_{j,1}$ with the semantic meaning "the value on f_j is large", $m_{j,2}$ with the semantic meaning "the value on f_j is medium", and $m_{j,3}$ with the semantic meaning "the value is closer to the small on f_j".

5.1. The facial salient characteristics analysis of male

In Figure 2, we show a segmental face image, whose gender is male. In order to distinguish each ethnic group, the symbol ξ_{c_1}, ξ_{c_2} and ξ_{c_3} is respectively to represents a set of semantic concepts of Mongolian, Korean and Hui.

According to the Algorithm 1, we can obtain the best optimal semantic concept set for each ethnic group, in which there are four different semantic concepts for ξ_{c_1}, four different semantic concepts for ξ_{c_2}, one semantic concepts for ξ_{c_3}. The semantic concept is detailed as follow:

ξ_{c_1} = "$m_{27,3} + m_{29,3}$" or "$m_{8,3} + m_{29,3}$" or "$m_{27,3} + m_{28,3}$";
ξ_{c_2} = "$m_{17,2}$";
ξ_{c_3} = "$m_{28,2} + m_{29,2}$" or "$m_{3,2} + m_{9,2}$" or "$m_{27,2} + m_{28,2}$" or "$m_{3,3} + m_{9,2} + m_{10,2} + m_{25,2} + m_{27,2} + m_{28,2} + m_{29,2}$";



Mongolian Korean Hui

Figure 2: A segmental ethnic face images of Mongolian, Korean and Hui. The gender is male

From semantic concepts, we can obviously gain facial features relating to the semantic concepts of each ethnic group. For example, the feature f_8, f_{27}, f_{28} and f_{29}, which correspond to $m_{8,3}$, $m_{27,3}$, $m_{28,3}$ and $m_{29,3}$ in the facial feature of Mongolian; the feature f_{17} corresponding to $m_{17,2}$ is significant for Korean; the features f_3, f_9, f_{10}, f_{25}, f_{27}, f_{28}, f_{29} are important for Hui as similar as that of Mongolian. Due to different measurement of all facial features, such as width, height, area and angle, we normalize the feature data in interval between 0 and 1. Then, we use the mean value and standard deviation to illustrate the distribution of all features in Figure 3.

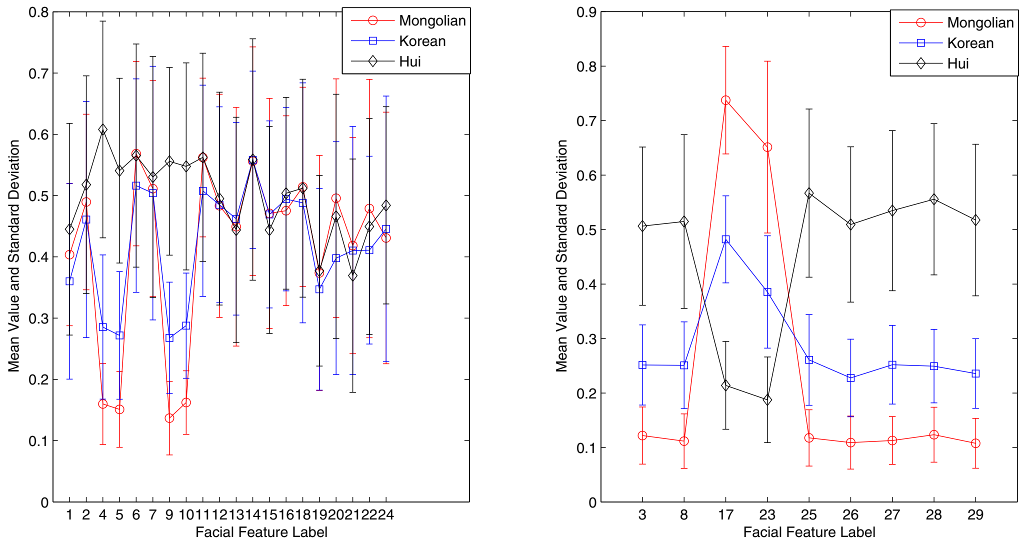


Figure 3: The mean value and standard deviation of facial features on the male of Mongolian, Korean and Hui.

From Figure 3, we can clearly observe the distributions of all facial features, which are used to describe Mongolian, Korean and Hui. Respectively, the standard deviation belonging to ethnic groups distributes in different region between 0 and 1, and each median value of standard deviation interval is the mean value for corresponding ethnic group in f_j. Therefore, we take the f_{17} as an example, the standard deviation respectively belonging to Mongolian, Korean and Hui distributes in three different regions, and each median value of each ethnic group is different from others. It is favorable for us to identify ethnic groups. Oppositely, because of approximate distribution of standard deviation, it is difficult to recognize the differences such as f_1, f_{13} and f_{14}. However, all the feature f_3, f_8, f_9, f_{10}, f_{17},

f_{25}, f_{27}, f_{28}, f_{29}, which are extracted by our method, are effectiveness for analyzing facial characteristics. Moreover, we verify the performance of features for describing ethnic facial character with K-means and FCM. The results are listed in Table 5 and Table 6.

Table 5: The cluster purity with the feature f_j for male based on K-means(%)

Ethnic	f_3			f_8			f_9			f_{10}			f_{17}		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	**71.55**	4.21	0.00	0.00	**74.36**	0.00	0.00	4.08	**70.94**	64.44	0.00	0.00	6.32	**94.19**	0.00
ξ_{c_2}	26.72	62.11	0.00	65.22	24.79	1.69	1.89	56.12	29.06	33.33	2.22	50.00	**86.32**	5.81	3.45
ξ_{c_3}	1.73	33.68	**100.0**	34.78	0.85	98.31	**98.11**	39.80	0.00	2.23	**97.78**	50.00	7.36	0.00	**96.55**

Ethnic	f_{25}			f_{27}			f_{28}			f_{29}			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	−−	−−	−−
ξ_{c_1}	0.00	0.00	66.92	67.97	0.00	0.00	**73.68**	0.00	3.53	**75.44**	1.08	0.00	−−	−−	−−
ξ_{c_2}	0.00	61.04	33.08	31.25	56.18	0.00	26.32	0.00	**70.59**	24.56	66.67	0.00	−−	−−	−−
ξ_{c_3}	**100.0**	38.96	0.00	0.78	43.82	**100.0**	0.00	**100.0**	25.88	0.00	32.25	**100.0**	−−	−−	−−

Table 6: The cluster purity with the feature f_j for male based on FCM(%)

Ethnic	f_3			f_8			f_9			f_{10}			f_{17}		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	**72.41**	4.21	0.00	0.00	**75.21**	0.00	0.00	3.22	68.55	66.17	0.00	0.00	8.25	**95.24**	0.00
ξ_{c_2}	26.72	62.11	0.00	67.03	23.94	1.67	1.96	54.84	30.65	31.58	2.08	54.02	**85.57**	4.76	3.45
ξ_{c_3}	0.87	33.68	**100.0**	32.97	0.85	**98.33**	**98.04**	41.94	0.80	2.25	**97.92**	45.98	6.18	0.00	**96.55**

Ethnic	f_{25}			f_{27}			f_{28}			f_{29}			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	−−	−−	−−
ξ_{c_1}	0.00	0.00	65.19	**70.16**	1.14	0.00	63.24	0.00	2.86	**75.65**	1.09	0.00	−−	−−	−−
ξ_{c_2}	0.00	54.32	34.07	29.03	61.36	0.00	36.03	0.00	58.57	24.35	67.39	0.00	−−	−−	−−
ξ_{c_3}	**100.0**	45.68	0.74	0.81	37.50	**100.0**	0.73	**100.0**	38.57	0.00	31.52	**100.0**	−−	−−	−−

According to test the performance of feature f_j, we use K-means and FCM to cluster the samples for three clusters, which is represented by $I_i (1 \leq i \leq 3)$. Then, we calculate the purity of each I_i, which is the percent of each ethnic group. Firstly, we set a threshold 70% to evaluate the consequence of feature f_j for describing the multi-ethnic facial characteristics. From the results in Table 5 and Table 6, we can clearly see that the purity of ξ_{c_1} is better with the feature f_8, f_{27}, f_{28}, f_{29} extracted by our method. Further, one can see from results of ξ_{c_1}, the feature f_{17} and f_3, which respectively belongs to ξ_{c_2} and ξ_{c_3}, have a better purity. It indicates that these features could explicitly distinguish face between ξ_{c_1} and ξ_{c_2} or ξ_{c_1} and ξ_{c_3}. As similar as the features belonging to ξ_{c_1}, the f_{17} is available for ξ_{c_2}, all the features of f_3, f_9, f_{10}, f_{25}, f_{27}, f_{28}, f_{29} are effective for ξ_{c_3}. The experiment verifies the angle features play an important role in facial ethnic characteristic such as f_3, f_8, f_9, f_{10}, f_{25}, f_{27}, f_{28}, f_{29}.

In addition, we have another test using the complex feature $\gamma_j (1 \leq j \leq 8)$. In according to semantic concepts of $\xi_{c_i} (1 \leq i \leq 3)$, the γ_i is respectively corresponded as follow: γ_1 represents "f_{27} and f_{29}", γ_2 represents "f_8 and f_{29}", γ_3 represents "f_{27} and f_{28}", γ_4 represents "f_{17}", γ_5 represents "f_{28} and f_{29}, γ_6 represents "f_3 and f_9", γ_7 represents "f_{27} and f_{28}" and γ_8 represents "f_3 and f_9 and f_{10} and f_{25} and f_{27} and f_{28} and f_{29}". γ_T is a feature set which includes all feature $f_j (1 \leq j \leq 29)$. The results tested by K-means and FCM are listed in Table 7 and Table 8.

The results show that the complex feature has a better purity than single feature f_j for γ_{c_i} in cluster I_i. For example, γ_1 includes two feature f_{27} and f_{29}. Comparing the results in Table 7 and Table 5, γ_1 has a better purity than that of belonging to f_{27} or f_{29} using K-means. Similarly, there is a contrast result between Table 8 and Table 6, the purity of γ_1 is greater than f_{27}, and that of is an approximate to f_{29} using FCM. The results of γ_1 is worse than other complex feature set.

From the results, we observe that the features including f_3, f_8, f_9, f_{10}, f_{17}, f_{25}, f_{27}, f_{28}, f_{29} play an important role in analyzing the male facial features of Mongolian, Korean and Hui. Especially, the angle feature is more significant than other features.

Table 7: The cluster purity with the feature γ_j for male based on K-means(%)

Ethnic	γ_1			γ_2			γ_3			γ_4			γ_5		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	**77.68**	0.00	1.10	**78.90**	0.00	69.60	0.00	0.00	6.32	**94.19**	0.00	0.00	**75.44**	1.08
ξ_{c_2}	0.00	22.32	**71.43**	**73.63**	21.10	0.00	30.40	0.00	60.47	**86.32**	5.81	3.45	0.00	24.56	66.67
ξ_{c_3}	**100.0**	0.00	28.57	25.27	0.00	**100.0**	0.00	**100.0**	39.53	7.36	0.00	**96.55**	**100.0**	0.00	32.25

Ethnic	γ_6			γ_7			γ_8			γ_T			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	68.50	0.00	0.00	69.60	0.00	0.00	1.12	0.00	**83.50**	32.27	19.37	65.23	−−	−−	−−
ξ_{c_2}	29.93	61.18	0.00	30.40	0.00	60.47	**82.02**	0.00	16.50	4.16	61.05	17.90	−−	−−	−−
ξ_{c_3}	1.57	38.82	**100.0**	0.00	**100.0**	39.53	16.86	**100.0**	0.00	63.57	19.58	16.87	−−	−−	−−

Table 8: The cluster purity with the feature γ_j for male based on FCM(%)

Ethnic	γ_1			γ_2			γ_3			γ_4			γ_5		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	**72.13**	0.00	1.10	**79.82**	0.00	**71.31**	0.00	1.14	8.25	**95.24**	0.00	0.00	**77.48**	2.13
ξ_{c_2}	0.00	27.87	65.88	62.39	20.18	0.00	28.69	0.00	62.50	**85.57**	4.76	3.45	0.00	22.52	69.15
ξ_{c_3}	**100.0**	0.00	34.12	36.51	0.00	100.0	0.00	**100.0**	36.36	6.18	0.00	**96.55**	**100.0**	0.00	28.72

Ethnic	γ_6			γ_7			γ_8			γ_T			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	**72.27**	2.35	0.00	**71.31**	0.00	1.14	1.18	0.00	**80.56**	33.27	16.32	64.47	−−	−−	−−
ξ_{c_2}	26.89	68.24	0.00	28.69	0.00	62.50	**81.18**	0.00	19.44	3.37	63.42	18.27	−−	−−	−−
ξ_{c_3}	0.84	29.41	**100.0**	0.00	**100.0**	36.36	17.65	**100.0**	0.00	63.36	20.26	17.26	−−	−−	−−

5.2. The facial salient characteristics analysis of female

In this section, we have an analysis for the female of Mongolian, Korean and Hui. Similarly, we assume ξ_{c_1}, ξ_{c_2} and ξ_{c_3} respectively represent a semantic concept set of Mongolian, Korean and Hui.

According to our optimization method, we would obtain the semantic concepts for three ethnic groups with the training set, and the result is detailed as following

ξ_{c_1} = "$m_{27,3}$+$m_{28,3}$" or "$m_{8,3}$+$m_{28,3}$" or "$m_{3,3}$+$m_{28,3}$" or "$m_{8,3}$+$m_{15,2}$+$m_{27,3}$";

ξ_{c_2} = "$m_{9,2}$+$m_{27,2}$" or "$m_{25,2}$+$m_{27,2}$" or "$m_{27,2}$" or "$m_{10,2}$+$m_{27,2}$";

ξ_{c_3} = "$m_{17,2}$";

Corresponding to the semantic concepts, we can easily match the facial features to multi-ethnic groups. ξ_{c_1} includes f_3, f_8, f_{15}, f_{27} and f_{28}. The f_9, f_{10}, f_{25} and f_{27} is belonged to ξ_{c_2}. The f_{17} plays an important role in describing

Mongolian Korean Hui

Figure 4: A Segmental ethnic face images of Mongolian, Korean and Hui. The gender is female.

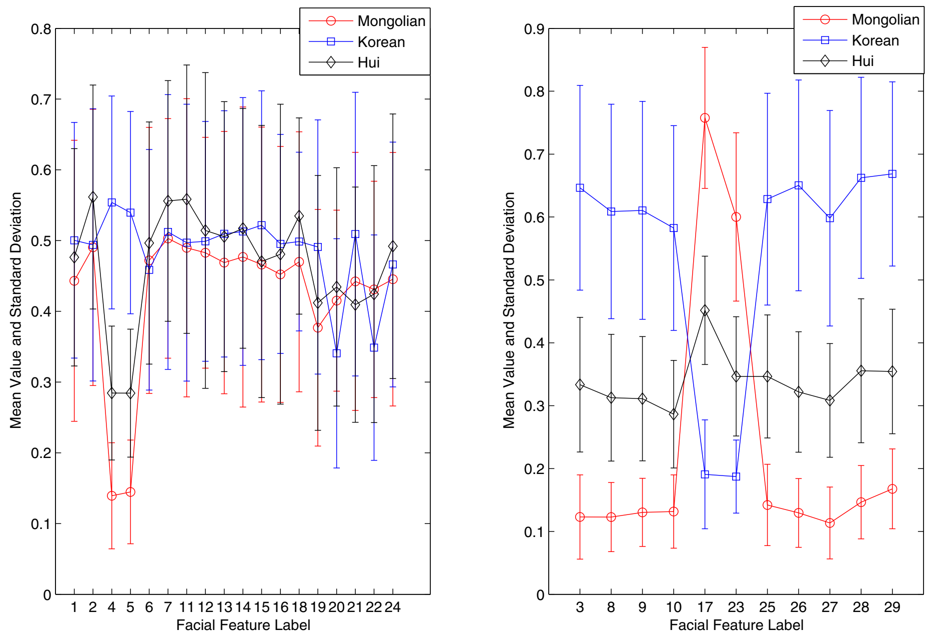


Figure 5: The mean value and standard deviation of facial features on the female of Mongolian, Korean and Hui.

ξ_{c_3}. Then, the mean value and standard deviation of all the features is shown for illustrating their differences in Figure 5.

As similar as Figure 3, the mean value and standard deviation of feature f_3, f_8, f_9, f_{10}, f_{15}, f_{17}, f_{25}, f_{27} and f_{28} are shown. Again, we verify the performance of features for describing ethnic facial characteristic with K-means and FCM. The results are listed in Table 9 and Table 10.

Table 9: The cluster purity with the feature f_j for female based on K-means(%)

Ethnic	f_3			f_8			f_9			f_{10}			f_{15}		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	1.02	**79.63**	**72.88**	0.00	0.98	0.00	0.00	61.70	0.00	0.00	57.24	36.76	21.62	36.80
ξ_{c_2}	**96.72**	31.63	0.00	1.69	**97.87**	41.18	**100.0**	55.81	1.42	**100.0**	64.10	1.97	23.53	41.89	34.40
ξ_{c_3}	3.28	67.35	20.37	25.43	2.13	57.84	0.00	44.19	36.88	0.00	35.90	40.79	39.71	36.49	28.80
Ethnic	f_{17}			f_{25}			f_{27}			f_{28}			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	−−	−−	−−
ξ_{c_1}	13.40	**98.67**	0.00	0.93	**71.67**	0.00	0.00	61.70	0.00	0.00	**76.99**	0.00	−−	−−	−−
ξ_{c_2}	3.09	0.00	**91.58**	46.73	0.83	**97.50**	**100.0**	2.13	59.79	31.82	0.00	**93.94**	−−	−−	−−
ξ_{c_3}	**83.51**	1.33	6.52	52.34	27.50	2.50	0.00	36.17	40.21	68.18	23.01	6.06	−−	−−	−−

In order to verify the influence for each ethnic group, we use K-means and FCM to cluster the samples for three clusters, which is represented by $I_i (1 \leq i \leq 3)$. Then, the purity of each I_i is shown in Table 9 and Table 10. We still set the threshold 70% to evaluate the performance of f_j. From the result, all the feature f_3, f_8, f_{17}, f_{25}, f_{27} and f_{28} have a better effect for ξ_{c_1}; the features f_3, f_8, f_9, f_{10}, f_{17}, f_{25}, f_{27} and f_{28} also have a good purity for ξ_{c_2}; the feature f_{17} is important for describing ξ_{c_3}; On the contrary, in according to the mean value and standard deviation, the performance of f_{15} is obvious weaker than other selected features.

However, certain features are not belonging to ξ_{c_i}, it still generates a excellent description for ξ_{c_i}. For example, feature f_{25} is used to describe ξ_{c_2}, but it has a similar effect for ξ_{c_1} . the reason is that not only f_{25} represents the facial characteristic, it expresses the difference among ethnic groups. Therefore, the f_{25} will provide a advantageous

Table 10: The cluster purity with the feature f_j for female based on FCM(%)

Ethnic	f_3			f_8			f_9			f_{10}			f_{15}		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	0.00	**76.32**	**74.14**	0.00	0.96	0.00	0.00	66.41	0.00	0.00	63.97	39.06	26.74	33.33
ξ_{c_2}	**98.21**	35.05	8.77	0.86	**97.87**	41.35	**97.78**	49.45	0.76	**100.0**	51.68	1.47	21.88	38.38	36.75
ξ_{c_3}	1.79	64.95	22.81	25.00	2.13	57.69	2.22	50.55	32.82	0.00	48.31	34.56	39.06	34.88	29.92
Ethnic	f_{17}			f_{25}			f_{27}			f_{28}			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	−−	−−	−−
ξ_{c_1}	13.00	**98.67**	0.00	0.93	**72.27**	0.00	0.00	**71.90**	0.00	0.00	**77.68**	0.00	−−	−−	−−
ξ_{c_2}	4.00	0.00	**93.48**	46.73	0.84	**95.12**	**98.00**	1.65	40.63	30.68	0.00	**94.03**	−−	−−	−−
ξ_{c_3}	**83.00**	1.33	6.52	52.34	26.89	4.88	2.00	26.45	59.37	69.32	22.32	5.97	−−	−−	−−

performance for ξ_{c_1}. In addition, the angle feature is still more important than others for distinguishing the facial characteristics.

Then, we test the purity using the complex feature $\gamma_j(1 \leq j \leq 9)$. Where, γ_1 represents "f_{27} and f_{28}", γ_2 represents "f_8 and f_{28}", γ_3 represents "f_3 and f_{28}", γ_4 represents "f_8 and f_{15} and f_{27}", γ_5 represents "f_9 and f_{27}", γ_6 represents "f_{25} and f_{27}", γ_7 represents "f_{27}" and γ_8 represents "f_{10} and f_{27}", γ_9 represents "f_{17}", γ_T is a feature set which includes all feature $f_j(1 \leq j \leq 29)$. The results tested by K-means and FCM are listed in Table 11 and Table 12.

Table 11: The cluster purity with the feature γ_j for female based on K-means(%)

Ethnic	γ_1			γ_2			γ_3			γ_4			γ_5		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	0.00	**71.90**	0.98	0.00	**72.88**	0.00	**79.63**	1.04	21.62	36.76	36.80	**74.36**	0.00	0.00
ξ_{c_2}	39.36	**98.08**	1.65	41.18	**97.87**	1.65	**95.24**	0.00	31.25	41.89	23.53	34.40	0.85	36.56	**96.49**
ξ_{c_3}	60.64	1.92	26.45	57.84	2.13	25.47	4.76	20.37	67.71	36.49	39.71	28.80	24.79	63.44	3.51
Ethnic	γ_6			γ_7			γ_8			γ_9			γ_T		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	**78.38**	0.00	0.00	0.00	**71.90**	0.00	0.00	**79.82**	0.00	13.00	**98.67**	0.00	66.96	15.28	16.93
ξ_{c_2}	0.00	**96.61**	34.02	**98.00**	1.65	40.63	**95.89**	1.83	21.18	4.00	0.00	**93.48**	11.08	66.47	18.63
ξ_{c_3}	21.62	3.39	65.98	2.00	26.45	59.37	4.11	18.35	**78.82**	**83.00**	1.33	6.52	21.96	18.25	64.44

Table 12: The cluster purity with the feature γ_j for female based on FCM(%)

Ethnic	γ_1			γ_2			γ_3			γ_4			γ_5		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	0.00	**73.73**	0.95	0.00	**74.78**	0.00	**79.63**	1.02	26.19	38.46	33.90	**75.65**	0.00	0.00
ξ_{c_2}	35.48	**98.21**	1.69	41.90	**97.87**	0.00	**96.72**	0.00	31.63	38.10	21.54	37.29	0.87	35.11	**96.55**
ξ_{c_3}	64.52	1.79	24.58	57.15	2.13	25.22	3.28	20.37	67.35	35.71	40.00	28.81	23.48	64.89	3.45
Ethnic	γ_6			γ_7			γ_8			γ_9			γ_T		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	**82.86**	0.00	0.00	0.00	**71.90**	0.00	0.00	**76.99**	0.00	13.00	**98.67**	0.00	67.50	17.70	18.70
ξ_{c_2}	0.00	**95.44**	28.13	**98.00**	1.65	40.63	**96.92**	1.77	28.09	4.00	0.00	**93.48**	11.04	64.44	19.94
ξ_{c_3}	17.14	4.76	**71.87**	2.00	26.45	59.37	3.08	21.24	**71.91**	**83.00**	1.33	6.52	21.46	17.86	61.36

According to the result, we can explicitly realize that the complex feature γ_i has better effect than f_j. The results of γ_1 is weaker than other complex feature set. For example, whether K-means or FCM, the f_{10} and f_{27} play a role for ξ_{c_2}. In addition, feature f_{27} also has a good description for ξ_{c_2} and ξ_{c_3} with FCM. However, γ_8 has a favorable effectiveness for ξ_{c_1}, ξ_{c_2} and ξ_{c_3}. It means that the angle feature is more important than others for expressing the facial differences among ethnic groups, and the complex angle feature is more significant than single that of. Meanwhile, we can demonstrate all the features, f_3, f_8, f_9, f_{10}, f_{15}, f_{17}, f_{25}, f_{27} and f_{28}, are meaningful for describing the female face of Mongolian, Korean and Hui.

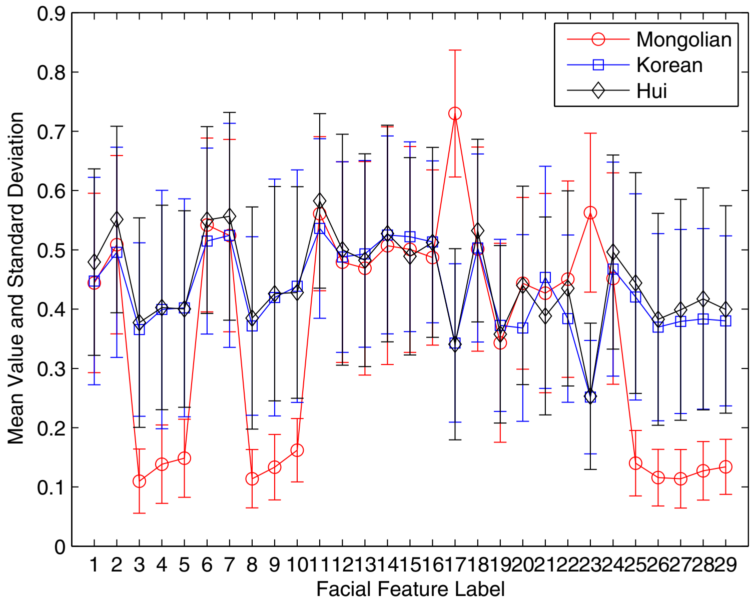


Figure 6: The mean value and standard deviation of facial features about Mongolian, Korean and Hui.

5.3. *The facial salient characteristics analysis for Mongolian, Korean and Hui*

When the gender is ignored, whether the facial features are similar in same ethnic group? In order to verify it, we reconstruct a mixed database according to ethnic label, in which each class includes male and female. Similarly, ξ_{c_1}, ξ_{c_2} and ξ_{c_3} still represent a set semantic concept set of Mongolian, Korean and Hui. We would still obtain the semantic concepts for three ethnic groups using our method. The semantic concepts will be illustrated as follows:

ξ_{c_1} = "$m_{9,3}$+$m_{27,3}$" or "$m_{25,3}$+$m_{26,3}$+$m_{28,3}$" or "$m_{8,3}$+$m_{28,3}$" or "$m_{9,3}$+$m_{25,3}$" or "$m_{8,3}$+$m_{28,3}$";

ξ_{c_2} = "$m_{4,2}$+$m_{18,2}$" or "$m_{4,2}$+$m_{5,2}$+$m_{28,2}$+$m_{29,2}$" or "$m_{2,2}$+$m_{7,3}$" or "$m_{4,2}$+$m_{17,2}$" or "$m_{4,2}$+$m_{13,2}$";

ξ_{c_3} = "$m_{1,2}$ + $m_{10,2}$" or "$m_{1,2}$ + $m_{8,2}$ + $m_{27,2}$" or "$m_{1,2}$ + $m_{29,2}$" or "$m_{1,2}$ + $m_{9,2}$ + $m_{28,2}$" or "$m_{1,2}$ + $m_{12,1}$";

Corresponding to the semantic concepts, there are many features selected for presenting the facial characteristics of multi-ethnic groups, which obtains f_1, f_2, f_4, f_5, f_7, f_8, f_9, f_{10}, f_{12}, f_{13}, f_{17}, f_{18}, f_{25}, f_{26}, f_{27}, f_{28} and f_{29}. Then, the mean value and standard deviation of all features are shown for illustrating their differences in Figure 6.

We can clearly realize the distribution of mean value and standard deviation usually distribute in similar region on any feature f_j so that it is difficult to differentiate face through these facial features, which can be observed in Figure 6. Further, for purpose of confirming whether the facial features could be represented by ignoring gender for each ethnic group. We have extracted the facial features with combined database, and respectively gained salient features of male and female. Then, the common features extracted from section 5.1, section 5.2 and section 5.3 including f_8, f_9, f_{10}, f_{17}, f_{25} f_{27} and f_{28}, will be chosen to verify the performance of combined database. The threshold is still set into 70%. If the purity of any cluster exceeds threshold with the feature f_j, it will be considered as a significant description for corresponding ethnic group.

From Table 13 and Table 14, we can clearly observe that the performance of the facial features is less than that of in Table 5, Table 6, Table 9 and Table 10. Moreover, it is difficult to differentiate facial characteristics between Korean and Hui. On the contrary, these facial features still conduct a better effectiveness for Mongolian, especially for the feature f_{17}. Meanwhile, in according to the results, we can acquire three results: (1) The facial characteristic of Mongolian is similar between male and female; (2) The facial characteristics of Korean and Hui are different between male and female; (3) The angle features play an important role in describing facial characteristics of Mongolian such as f_8, f_9, f_{25} and f_{28}.

Table 13: The cluster purity with the feature f_j based on K-means(%)

Ethnic	f_8			f_9			f_{10}			f_{17}		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	2.03	0.00	**72.77**	**71.13**	0.00	2.66	0.00	68.25	1.68	0.00	**95.36**	15.05
ξ_{c_2}	55.33	41.75	11.91	15.06	51.85	46.81	54.81	17.46	44.13	49.44	0.67	44.17
ξ_{c_3}	42.64	58.25	15.32	13.81	48.15	50.53	45.19	14.29	54.19	50.56	3.97	40.78
Ethnic	f_{25}			f_{27}			f_{28}			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	−−	−−	−−
ξ_{c_1}	0.00	0.52	**70.73**	0.53	68.50	0.00	0.00	69.35	1.85	−−	−−	−−
ξ_{c_2}	39.80	53.40	15.85	57.45	13.78	39.78	42.40	14.52	56.17	−−	−−	−−
ξ_{c_3}	60.20	46.08	13.42	42.02	17.72	60.22	57.60	16.13	41.98	−−	−−	−−

Table 14: The cluster purity with the feature f_j based on FCM(%)

Ethnic	f_8			f_9			f_{10}			f_{17}		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
ξ_{c_1}	0.00	0.00	**70.56**	62.14	0.00	0.59	0.00	63.14	1.20	0.00	**94.27**	13.64
ξ_{c_2}	58.20	40.82	12.10	19.64	49.41	48.82	53.19	18.98	46.71	49.44	0.64	45.45
ξ_{c_3}	41.80	59.18	17.34	18.22	50.59	50.59	46.81	17.88	52.09	50.56	5.09	40.91
Ethnic	f_{25}			f_{27}			f_{28}			−−		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3	−−	−−	−−
ξ_{c_1}	0.00	0.55	67.44	0.00	65.06	0.00	0.00	**75.44**	1.81	−−	−−	−−
ξ_{c_2}	41.49	53.00	17.05	57.92	15.61	38.55	43.26	11.40	56.02	−−	−−	−−
ξ_{c_3}	58.51	46.45	15.51	42.08	19.33	61.45	56.74	13.16	42.17	−−	−−	−−

5.4. The facial salient characteristics analysis for female of Brazilian and American

The FEI database includes Brazilian, who originates from Portuguese. CK+ mainly contains Caucasian-American. We make a combined data, which includes Brazilian and American, to interpret the typical characteristic of different ethnic groups. Respectively, we choose 165 face images(FEI 100, CK+ 65) of frontal position and neutral expression.

Similarly, we assume ξ_{c_1} and ξ_{c_2} respectively represents a set semantic concept set of Brazilian and American. The facial semantic concepts of female as follow:

ξ_{c_1} = "$m_{17,3}$" or "$m_{3,1}m_{8,1}$" or "$m_{8,1}m_{9,1}$" or "$m_{9,1}m_{25,1}$" ;

ξ_{c_2} = "$m_{17,1}$" or "$m_{3,3}m_{17,1}$" or "$m_{8,3}m_{17,1}$";

In according to the semantic concepts, we can easily gain the features of each ethnic group that the features including f_3, f_8, f_9, f_{17} and f_{25} are belonged to ξ_{c_1}, and the feature f_3, f_8 and f_{17} are used to describing ξ_{c_2}. Then, we still have an experiment on mean value and standard deviation of these features by extracted. The result is shown in Figure 8.

From the Figure 8, the mean value and standard deviation of all features are illustrated. Although, their distributions are approximate, we still test the performance of features chosen by our method with K-means and FCM. The data is formed two clusters, $I_i(1 \leq i \leq 2)$ is respectively represented Brazilian and American. In addition, the number of Brazilian is larger than American. The results are listed in Table 15 and Table 16.

Table 15: The cluster purity with the feature f_j for female(Brazilian-American) based on K-means(%)

Ethnic	f_3		f_8		f_9		f_{17}		f_{25}	
	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2
ξ_{c_1}	41.94	**84.72**	**83.78**	41.76	25.00	**85.19**	23.19	**80.33**	**81.93**	39.02
ξ_{c_2}	58.06	15.28	16.22	58.24	**75.00**	14.81	**76.81**	19.67	18.07	60.98

We can clearly observe the angle features also play an important role on representing facial characteristics of Brazilian and American. Then, we test the purity using the complex feature $\gamma_j(1 \leq j \leq 6)$. In which, γ_1 represents "f_{17}", γ_2 represents "f_3 and f_8", γ_3 represents "f_8 and f_9", γ_4 represents "f_3 and f_{17}", γ_5 represents "f_8 and f_{17}", γ_6 represents "f_9 and f_{25}", γ_T is a feature set which includes all feature $f_j(1 \leq j \leq 29)$.

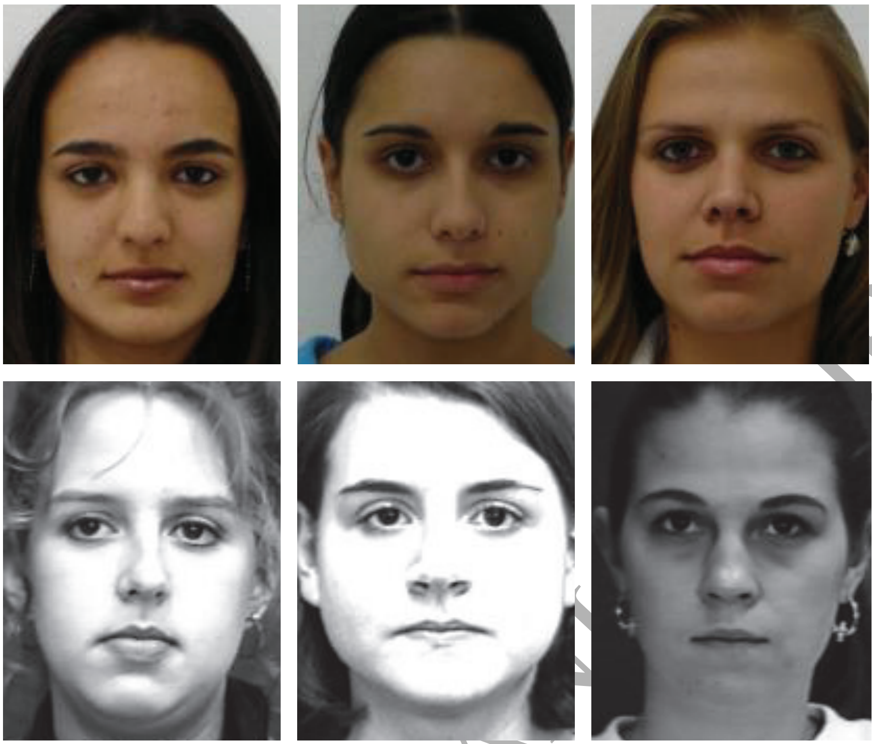


Figure 7: A segmental face images belonging to FEI and CK+ database. The gender of all face is female. The face image is belonged to FEI in fist row, and that of is belonged to CK+ in second row

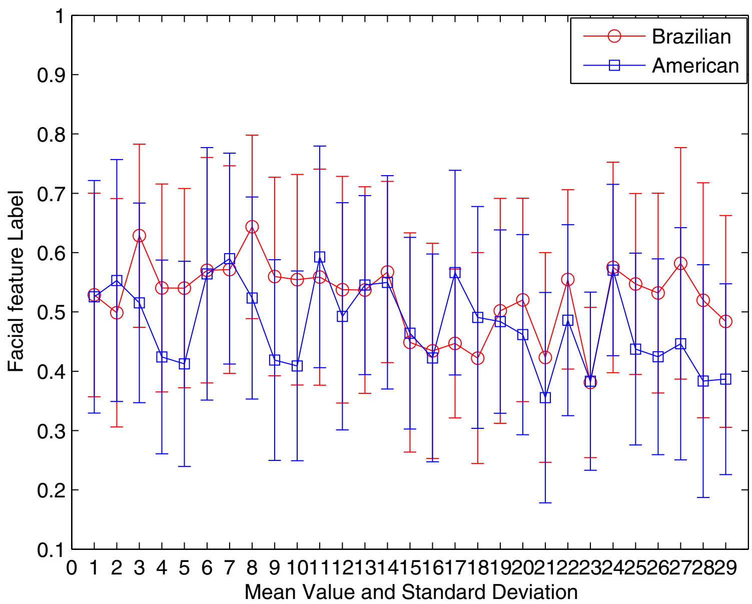


Figure 8: The mean value and standard deviation of facial features on the female of Brazilian and American

Table 16: The cluster purity with the feature f_j for female(Brazilian-American) based on FCM(%)

Ethnic	f_3		f_8		f_9		f_{17}		f_{25}	
	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2
ξ_{c_1}	44.87	**74.71**	**74.71**	44.87	24.32	**83.93**	24.64	**80.33**	**74.71**	44.87
ξ_{c_2}	55.13	25.29	25.29	55.13	**75.68**	16.07	**75.36**	19.67	25.29	55.13

Table 17: The cluster purity with the feature γ_j for female(Brazilian-American) based on K-means(%)

Ethnic	γ_1		γ_2		γ_3		γ_4		γ_5		γ_6		γ_T	
	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2
ξ_{c_1}	23.19	**80.33**	41.76	**83.78**	**91.67**	36.56	32.94	**90.00**	32.53	**89.02**	**85.90**	37.93	65.56	54.67
ξ_{c_2}	**76.81**	19.67	58.24	16.22	8.33	63.44	67.06	10.00	67.47	10.98	14.10	62.07	34.44	45.33

According to the Table 17 and Table 18, the results again certify that the purity of the complex feature $\gamma_j(1 \leq j \leq 6)$ is better than that of γ_T. Specially, the feature subsets including f_3, f_8, f_9, f_{17} and f_{25} have a better effectiveness for describing differences of facial characteristics between Brazilian and American. Meanwhile, the results denotes the angle features play an important role for describing facial characteristics among various ethnic groups.

In summary, we gain some achievements as follow.(1)Firstly, in order to AFS theory, we design a new approach to extract facial features according to its distribution. There are two advantages: one is that the previous model is not necessary, anther is the features can be easily comprehended by semantic concept. (2)Secondly, on the basis of gender discrepancy, the facial features chosen by our method are beneficial to differentiate multi-ethnic groups. (3) Finally, considering the ethnic label, the common features of male and female are extracted respectively. The result indicates the facial features of Mongolian are similar between male and female, and confusable between Korean and Hui implying frequently communication. Moreover, we have a significance discover that the angle features have better performance than other distance features, such as f_3 "the angle between left eyes canthus and left eyes top point", f_8 "the angle of between right canthus and right eye top point", f_9 "the angle between right eyes top and down point and right eyes inner canthus", f_{25} "the angle between zygomatic and underjaw apex".

5.5. Comparative analysis

In this section, we address the comparison of performance between our method and other feature selection methods such as Boosting[45], information gain(InfGain)[46], information gain ratio(GainRatio)[47]. Using the features extracted by feature selection methods, the classification testing will be executed by some state-of-the-art classifiers such as PCA[3], C4.5 [48], Bayes[49], DecisionTable [50] and Cart [51]. The results are illustrated in Table19, Table20 and Table21.

Generally, our method has a better advantage than other methods. In table 19, the classification accuracy using the features extracted by our method is better than other selection methods based on the classifiers such as PCA, Bayes and Cart. And the result also approximates GainRatio and Boosting based on C4.5 and DecisionTable. According to Table 20, it denotes the classification results are better than other method using the features extracted by our method on PCA, Bayes, DecisionTable and Cart. Similarly, in the public database(see Table 21), the features extracted by our method still has a better effectiveness on PCA, Bayes, DecisionTable and Cart.

Table 18: The cluster purity with the feature γ_j for female based on FCM(%)

Ethnic	γ_1		γ_2		γ_3		γ_4		γ_5		γ_6		γ_T	
	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2	I_1	I_2
ξ_{c_1}	24.64	**80.33**	44.87	**74.71**	**74.71**	44.87	44.87	**74.71**	44.87	**74.71**	**74.71**	44.87	66.67	53.85
ξ_{c_2}	**75.36**	19.67	55.13	25.29	25.29	55.13	55.13	25.29	55.13	25.29	25.29	55.13	33.33	46.15

Table 19: The comparison of accuracy rate among classifiers based on CEFD(male)(%)

Methods \ Classifiers	PCA	C4.5	Bayes	DecisionTable	Cart
Boosting	89.26	90.00	92.59	**89.63**	90.37
InfGain	91.11	89.26	89.26	89.26	88.89
GainRatio	90.74	**90.74**	92.22	88.89	90.00
Our method	**92.22**	90.37	**93.33**	89.26	**92.59**

Table 20: The comparison of accuracy rate among classifiers based on CEFD(female)(%)

Methods \ Classifiers	PCA	C4.5	Bayes	DecisionTable	Cart
Boosting	90.37	90.37	90.74	90.74	89.63
InfGain	90.74	**91.48**	89.63	92.22	88.89
GainRatio	88.89	88.89	90.74	**92.59**	91.48
Our method	**91.48**	89.29	**92.22**	**92.59**	**91.85**

6. Conclusion

In this paper, we propose a novel method to represent the differences of multi-ethnic facial appearance, which use the semantic concept extracted from the facial features based on AFS. There are two advantages, the one is that it bridges the gap between primitive features and semantic concepts, which means each the chosen feature will be corresponding to a semantic concept, which is conveniently comprehended; another is that it also presents a new approach to reduce dimensionality for extracting salient facial features. Then, we use analyze the Chinese multi-ethnic facial characteristic. Firstly, landmark method is applied to automatically detect the facial components using landmarks on each face image. Secondly, the semantic concept is built for expressing each facial component such as eye width, pupils distance, nose width and other facial features, which can vividly describe facial characteristics. Thirdly, we operate the simple semantic concept based on AFS framework for gaining the complex concepts, and construct an optimal criterion to pick out salient semantic concept sets from the complex semantic concept sets to reveal the facial differences of ethnic groups in detail.

After the facial features of all ethnic groups are extracted, we analyze the facial differences and similarities among Mongolian, Korean and Hui using K-means and FCM. We discover that the gender has an important effect for various ethnic groups on facial characteristics, and harvest other research achievement about facial features. On the basis of gender, we respectively extract the facial features, which can clearly represent facial characteristics among the ethnic groups. For example, whether male or male, it can easily distinguish multi-ethnic groups using the features such as f_8, f_9, f_{17}, f_{25}, f_{28}. When the gender is ignored, Korean and Hui become difficult to be identified, but Mongolian is still conveniently recognized from ethnic groups. The results propose a significant presentation when we identify multi-ethnic face, some ethnic groups need two sets of facial features to differentiate male or female such as Korean and Hui, but others only use one set of that because of the similarities between male and female as Mongolian. Meanwhile, experiments show that the angle features have an important role in facial characteristics.

Finally, in order to demonstrate the performance of our method, we conduct the experiment on the combined database using FEI and CK+. Although, there are fusion between Brazilian-Portuguese and Caucasian-American because of similarity culture, custom and etc., the similar features will appear on their face. Furthermore, the experi-

Table 21: The comparison of accuracy rate among classifiers based on Brazilian-American(female)(%)

Methods \ Classifiers	PCA	C4.5	Bayes	DecisionTable	Cart
Boosting	69.70	81.82	76.37	81.82	81.82
InfGain	70.30	81.82	81.21	82.43	83.03
GainRatio	**70.91**	**84.24**	78.79	81.21	83.03
Our method	**70.91**	83.03	**81.82**	**83.03**	**84.24**

ments verify the viability and reliability of our method. Meanwhile, it denotes that the angle features are key features for describing facial characteristics of various ethnic groups.

7. Acknowledge

This work is supported by Natural Science Foundation of China(No.61370146) and Liaoning Science & Technology of Liaoning Province of China(No.2013405003).

References

[1] H. Zhao, P. C. Yuen, J. T. Kwok, A novel incremental principal component analysis and its application for face recognition, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 36 (2006) 873–886.
[2] J. Yang, D. Zhang, J. Yang, Constructing pca baseline algorithms to reevaluate ica-based face-recognition performance, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 37 (2007) 1015–1021.
[3] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of cognitive neuroscience 3 (1991) 71–86.
[4] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, Pattern Analysis and Machine Intelligence, IEEE Transactions on 19 (1997) 711–720.
[5] M. S. Bartlett, J. R. Movellan, T. J. Sejnowski, Face recognition by independent component analysis, Neural Networks, IEEE Transactions on 13 (2002) 1450–1464.
[6] W. Chen, M. J. Er, S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 36 (2006) 458–466.
[7] X. Jing, D. Zhang, A face and palmprint recognition approach based on discriminant dct feature extraction, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 34 (2004) 2405–2415.
[8] X. Geng, K. Smith Miles, Z. Zhou, L. Wang, Face image modeling by multilinear subspace analysis with missing values, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 41 (2011) 881–892.
[9] Y. Xu, X. Li, J. Yang, Z. Lai, D. Zhang, Integrating conventional and inverse representation for face recognition, Cybernetics, IEEE Transactions on 44 (2014) 1738–1746.
[10] C. Tian, G. Fan, X. Gao, Q. Tian, Multiview face recognition: from tensorface to v-tensorface and k-tensorface, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (2012) 320–333.
[11] J. Lu, Y. P. Tan, Regularized locality preserving projections and its extensions for face recognition, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 40 (2010) 958–963.
[12] S. Du, R. K. Ward, Improved face representation by nonuniform multilevel selection of gabor convolution features, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 39 (2009) 1408–1419.
[13] S. H. Lee, J. Y. Choi, Y. M. Ro, K. N. Plataniotis, Local color vector binary patterns from multichannel face images for face recognition, Image Processing, IEEE Transactions on 21 (2012) 2347–2353.
[14] D. Huang, H. Ding, C. Wang, Y. Wang, G. Zhang, L. Chen, Local circular patterns for multi-modal facial gender and ethnicity classification, Image and Vision Computing 32 (2014) 1181–1193.
[15] F. A. Pujol, J. C. García, Computing the principal local binary patterns for face recognition using data mining tools, Expert Systems with Applications 39 (2012) 7165–7172.
[16] F. Dornaika, A. Bosaghzadeh, Exponential local discriminant embedding and its application to face recognition, Cybernetics, IEEE Transactions on 43 (2013) 921–934.
[17] M. Liu, S. Li, S. Shan, X. Chen, Au-inspired deep networks for facial expression feature learning, Neurocomputing 159 (2015) 126–136.
[18] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, L. Prevost, Facial action recognition combining heterogeneous features via multikernel learning, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (2012) 993–1005.
[19] R. S. Smith, T. Windeatt, Facial action unit recognition using multi-class classification, Neurocomputing 150 (2015) 440–448.
[20] Y. Li, J. Chen, Y. Zhao, Q. Ji, Data-free prior model for facial action unit recognition, Affective Computing, IEEE Transactions on 4 (2013) 127–141.
[21] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar, Describable visual attributes for face verification and image search, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (2011) 1962–1977.
[22] Y. Tong, W. Liao, Q. Ji, Facial action unit recognition by exploiting their dynamic and semantic relationships, IEEE Transactions on Pattern Analysis & Machine Intelligence (2007) 1683–1699.
[23] B. C. Chen, Y. Chen, Y. H. Kuo, W. H. Hsu, Scalable face image retrieval using attribute-enhanced sparse codewords, Multimedia, IEEE Transactions on 15 (2013) 1163–1173.
[24] R. Brunelli, T. Poggio, Face recognition: Features versus templates, IEEE Transactions on Pattern Analysis & Machine Intelligence (1993) 1042–1052.
[25] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active shape models-their training and application, Computer vision and image understanding 61 (1995) 38–59.
[26] G. J. Edwards, T. F. Cootes, C. J. Taylor, Face recognition using active appearance models, in: Computer VisionECCV98, Springer, 1998, pp. 581–595.
[27] T. Lee, S.-K. Park, M. Park, An effective method for detecting facial features and face in human–robot interaction, Information sciences 176 (2006) 3166–3189.

[28] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, I. Matthews, S. Sridharan, In the pursuit of effective affective computing: The relationship between features and registration, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (2012) 1006–1016.

[29] J. Yi, X. Mao, L. Chen, Y. Xue, A. Compare, Facial expression recognition considering individual differences in facial structure and texture, Computer Vision, IET 8 (2014) 429–440.

[30] Z. Cui, W. Li, D. Xu, S. Shan, X. Chen, Fusing robust face region descriptors via multiple metric learning for face recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3554–3561.

[31] P. Conilione, D. Wang, Fuzzy approach for semantic face image retrieval, The Computer Journal (2012) bxs041.

[32] Y. Ren, Q. Li, W. Liu, L. Li, Semantic facial descriptor extraction via axiomatic fuzzy set, Neurocomputing 171 (2016) 1462–1474.

[33] X. Duan, C. Wang, X. Liu, Z. Li, j. Wu, H. Zhang, Ethnic features extraction and recognition of human faces, in: Advanced Computer Control (ICACC), 2010 2nd International Conference on, volume 2, IEEE, pp. 125–130.

[34] X. Liu, W. Wang, T. Chai, The fuzzy clustering analysis based on afs theory, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 35 (2005) 1013–1027.

[35] X. Liu, The fuzzy sets and systems based on afs. structure, ei algebra and eii algebra, Fuzzy Sets and Systems 95 (1998) 179–188.

[36] X. Liu, W. Pedrycz, Axiomatic fuzzy set theory and its applications, volume 244, Springer, 2009.

[37] H. Xiong, J. Wu, J. Chen, K-means clustering versus validation measures: a data-distribution perspective, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 39 (2009) 318–331.

[38] K. Krishna, M. N. Murty, Genetic k-means algorithm, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 29 (1999) 433–439.

[39] A. Stetco, X. Zeng, J. Keane, Fuzzy c-means++: fuzzy c-means with effective seeding initialization, Expert Systems with Applications 42 (2015) 7541–7548.

[40] L. Chen, C. P. Chen, M. Lu, A multiple-kernel fuzzy c-means algorithm for image segmentation, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 41 (2011) 1263–1274.

[41] C. E. Thomaz, G. A. Giraldi, A new ranking method for principal components analysis and its application to face image analysis, Image and Vision Computing 28 (2010) 902–913.

[42] E. Z. Tenorio, C. E. Thomaz, Análise multilinear discriminante de formas frontais de imagens 2d de face, in: Proceedings of the X Simposio Brasileiro de Automacao Inteligente SBAI, pp. 266–271.

[43] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, IEEE, pp. 94–101.

[44] T. Kanade, J. F. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, IEEE, pp. 46–53.

[45] A. J. Grove, D. Schuurmans, Boosting in the limit: Maximizing the margin of learned ensembles, in: AAAI/IAAI, pp. 692–699.

[46] H. Uğuz, A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm, Knowledge-Based Systems 24 (2011) 1024–1032.

[47] J. Dai, Q. Xu, Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification, Applied Soft Computing 13 (2013) 211–221.

[48] J. R. Quinlan, C4. 5: programs for machine learning, Elsevier, 2014.

[49] M. Ramoni, P. Sebastiani, Robust bayes classifiers, Artificial Intelligence 125 (2001) 209–226.

[50] J. Huysmans, K. Dejaeger, C. Mues, J. Vanthienen, B. Baesens, An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models, Decision Support Systems 51 (2011) 141–154.

[51] T.-S. Lee, C.-C. Chiu, Y.-C. Chou, C.-J. Lu, Mining the customer credit using classification and regression tree and multivariate adaptive regression splines, Computational Statistics & Data Analysis 50 (2006) 1113–1130.

Zedong Li received the B.S.c. degree in Automation form Chongqing University of Technology, Chongqing, China in 2006, the M.S.c. degree in Automation from Shenyang Ligong University, Shenyang, China, in 2011. Currently, he is a PhD candidate in the Institute of System Science, Northeastern University. His research interests include fuzzy system, knowledge discovery, and data process.

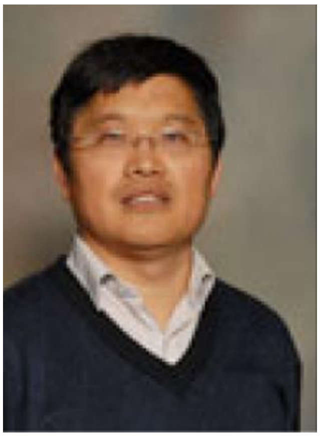
Xiaodong Duan is the director of Dalian Key Lab of Digital Technology for National Culture, Dalian Nationalities University, Dalian, PR. China. He received the the B.S. degrees in computer science and technology from Nankai University, Tianjin, P. R. China in 1985, and received the M. S. and Ph.D. degrees in applied mathematics and computer software and theory from Northeastern University, Shenyang, P. R. China in 1988 and 2001, respectively. His research interests include pattern recognition and data mining.

Qingling Zhang received the B.Sc. and M.Sc. degrees from the Mathematics

Department and the Ph.D. degree from the Automatic Control Department of Northeastern University, China, in 1982, 1986 and 1995, respectively. Since1997, he has been a Professor with Northeastern University. He is also a member of the University Teaching Advisory Committee of National Ministry of Education. He has published six books and more than 230 papers about control theory and applications. Dr. Zhang received 14 prizes from central and local governments for his research. He has also received the Golden Scholarship from Australia in 2000. His current research interests cover fuzzy system, singular systems, data-driven control, and robust control

Cunrui Wang received the B.S. degrees in computer science and technology from Dalian Nationalities University, Dalian, P. R. China, in 2002, and received the M.S. degrees in computer software and theory from Northeastern University, Shenyang, P. R. China, in 2005. Currently, he is a PhD candidate in the Institute of System Science, Northeastern University. His research interests include biometrics identification and data mining.

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, P. R. China, in 2012 and 2014, respectively. Currently, he is a PhD candidate in the Research Center of Information and Control, Dalian University of Technology. His research interests include machine learning and data mining.



Wanquan Liu received the B.S.c degree in Applied Mathematics from Qufu Normal University, PR China, in1985, the MSc degree in Control Theory and Operation Research from Chinese Academy of Science in 1988, and the PhD degree in Electrical Engineering from Shanghai Jiaotong University, in1993. He once held the ARC Fellowship, U2000 Fellowship and JSPS Fellowship and attracted research funds from different resources over 2million dollars. He is currently an associate professor in the Department of Computing at Curtin University and is in editorial board forseven international journals. His current research interests include large-scale pattern recognition, signal processing, machine learning, and control systems.