Received: 24 April 2018 | Revised: 26 June 2018 | Accepted: 5 July 2018

DOI: 10.1002/widm.1278

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

FOCUS ARTICLE

Facial feature discovery for ethnicity recognition

Cunrui Wang[1,2] | Qingling Zhang[2] | Wanquan Liu | Yu Liu[1] | Lixin Miao[1]

The study was conducted with the approval of Dalian University and written, informed consent was obtained from each study participant who understood their photographs would be used for non-profit scientific research. [Correction added on 02 August 2019 after first online publication: the preceding statement has been added to clarify some issues regarding the participants of the study.]

[1]Dalian Key Lab of Digital Technology for National Culture & Institute of System Science, Northeastern University, Dalian Nationalities University, Dalian, China

[2]Institute of System Science, Northeastern University, Shenyang, China

Correspondence
Cunrui Wang, Dalian Key Lab of Digital Technology for National Culture & Institute of System Science, Northeastern University, Dalian Nationalities University, China.
Email: cunruiwang@qq.com

Funding information
National Natural Science Foundation of China, Grant Number: 61562093, 61772575; China Education & Research Network Innovation Project, Grant Numbers: NGII20170419, NGII20170631

[Correction added on 10 January 2022, after first online publication: The Publisher has removed a co-author's affiliation to Curtin University. For full details please see https://doi.org/10.1002/widm.1446.]

The salient facial feature discovery is one of the important research tasks in ethnical group face recognition. In this paper, we first construct an ethnical group face dataset including Chinese Uyghur, Tibetan, and Korean. Then, we show that the effective sparse sensing approach to general face recognition is not working anymore for ethnical group facial recognition if the features based on whole face image are used. This is partially due to a fact that each ethnical group may have its own characteristics manifesting only in specified face regions. Therefore, we will analyze the particularity of three ethnical groups and aim to find the common characterizations in some local regions for the three ethnical groups. For this purpose, we first use the facial landmark detector STASM to find some important landmarks in a face image, then, we use the well-known data mining technique, the mRMR algorithm, to select the salient geometric length features based on all possible lines connected by any two landmarks. Second, based on these selected salient features, we construct three "T" regions in a face image for ethnical feature representation and prove them to be effective areas for ethnicity recognition. Finally, some extensive experiments are conducted and the results reveal that the proposed "T" regions with extracted features are quite effective for ethnical group facial recognition when the L_2-norm is adopted using the sparse sensing approach. In comparison to face recognition, the proposed three "T" regions are evaluated on the olivetti research laboratory face dataset, and the results show that the constructed "T" regions for ethnicity recognition are not suitable for general face recognition.

This article is categorized under:
- Algorithmic Development > Structure Discovery
- Algorithmic Development > Biological Data Mining
- Fundamental Concepts of Data and Knowledge > Knowledge Representation
- Technologies > Classification

KEYWORDS
facial ethnic feature, feature discovery, sparse coding, sparse representation

1 | INTRODUCTION

The analysis of race, nation, and ethnical groups based on facial images is a popular topic recently in face recognition community (Fu, He, & Hou, 2014). With rapid advance of people globalization, face recognition has great application potential in border control, customs check, and public security. Meanwhile, it is also an important research branch in physical anthropology. Usually, facial features are influenced by gene, environment, society, and other factors comprehensively. However, the gene of one ethnical group is hardly unique and it may include various gene fragments from some other ethnical groups. Hence, it may lead to the similarities of facial features among several ethnicities (Jianwen, Lihua, Lilongguang, & Shourong, 2010). Therefore, it is significant to analyze facial attributes for different ethnicities in computationally artificial intelligence. This work is also helpful to the research in anthropology as it may indicate the facial features evolution (Cunrui, Qingling, Xiaodong, Yuangang, & Zedong, 2018).

This paper focuses on the analysis of some Chinese ethnical groups. First, it is necessary for us to differentiate three definitions, namely race, nation, and ethnicity (Wade, 2007). Race is a concept which is formed based on the differences from physical structures such as skin, hair, and and so on, while nation is a social-oriented concept which refers to a community based on economics, language, and culture of a given area. Ethnicity describes a group of people, who have similar gene, culture, and language in geologically close regions. One can find that race and ethnicity are closely related though they have differences. For example, Chinese includes ethnical groups such as Han, Korean, Jing, Mongolian, Tibetan, Qiang, Miao, Turkic, Jurchen, and so on (Shiyuan, 2002). Based on homologous gene, ethnicities are steady groups and their facial features are regular and exhibit certain patterns. Although race and ethnicity have close relationship, the analysis of facial features among ethnicities is more difficult than that of race as the discrimination of facial features from different ethnicities is more difficult than that from different races (Fu et al., 2014).

Also, in cognitive process for a human face, human brains receive ethnicity or race information prior to age, gender, and expression. As shown in Figure 1, the information of ethnicity or race is processed in 80–120 ms, and the rest features, such as age and gender, are then gradually perceived later (Ito & Bartholow, 2009). This implies that race or ethnicity information is very important in face recognition.

In recent years, the sparse representation (SR) has broad applications in face recognition, expression recognition, and age estimation (Ortiz, Wright, & Shah, 2013; Ptucha, Tsagkatakis, & Savakis, 2011; Sun, Wang, & Tang, 2015; Wagner et al., 2012), but rarely used in ethnical group facial analysis (Fu et al., 2014). Although the SR has high effectiveness in general for face recognition, it is not effective in ethnicity recognition with the features from the whole face image, as demonstrated in this paper, especially when the sample size of each ethnicity is small. We believe this phenomena is due to a fact that the significant facial features for each ethnicity are only located in some typical regions on a face image and the features from other regions will reduce the discriminative capability for ethnical group recognition. Thus, we need to find some salient regions for these corresponding features and discover the effective facial features for ethnicity recognition.

2 | PRELIMINARIES

The past decade has witnessed the increasing popularity of facial ethnical recognition. Many researches have been conducted for extracting ethnical facial features using various approaches such as geometrical feature, holistic feature, local feature, and fusion features. Chan and Bledsoe (1965) analyzed the facial features of the White by using the distance and ratio of facial geometrical features. According to geometrical relationship of eyes, mouth, and underjaw, Kanade (1977) matched face images in a dataset constructed by himself. Brunelli and Poggio (1993) measured face similarity using facial geometrical features, which include nose length, mouth width, and underjaw shape, and the results indicated that geometrical features could be used to identify ethnical groups quite well. Brooks and Gwinn (2010) analyzed the differences between the White and Black using the skin color. According to their proposed skin color model, Gwinn extracted the facial features from Asian and European. Akbari and Mozaffari (Mar. 2012) explored the relations of facial skin color using south Indian, Australian, and African. Anzures, Pascalis, Quinn, Slater, and Lee (2011) confirmed that skin color was very sensitive to illumination, so that the skin color was usually fused in combined features to classify people primarily. Since Turk and Pentland (1991) proposed principal component analysis (PCA) in facial feature analysis including eyes, nose, and mouth successfully, PCA has been a popular method in face recognition. Based on PCA, Levine (1996) conducted facial feature extraction between Burman and non-Burman. Awwad, Ahmad, and Salameh (2013) accomplished facial features analysis for Arabian, Asian, and Caucasian. Based on scale, illumination and pose, Yan and Zhang (2009) used PCA to analyze the facial features on CMU and UCSD databases. Recently, many deep neural network methods are also used for face analysis and recognition (Chen, Zhang, Dong, Le, & Rao, 2017; Luan et al., 2018; Trigeorgis, Snape, Kokkinos, & Zafeiriou, 2017; Zhang, Song, & Qi, 2017). Srinivas et al. (2017) focused on predicting ethnicity using a convolutional neural network (CNN) with the Wild East Asian Face Dataset.

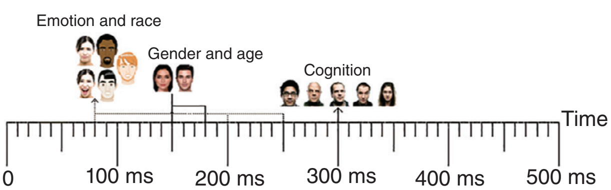

FIGURE 1 The order of attribute identification in face recognition

Local features can reduce the influence of illumination and obstacle occlusion, which are usually performing better than holistic features. For example, wavelet and local binary pattern (LBP) had shown their effectiveness on FERET database (Kumar, Berg, Belhumeur, & Nayar, 2011; Salah, Du, & Al-Jawad, 2013). In addition, Fu, Yang, and Hou (2011) analyzed facial expression using embedded topographic independent component analysis (TICA), the results showed the advantages of local features. However, the combined features which usually include skin color features, local wavelets features, and holistic features were used in practice instead of a single type of facial features. Ding, Huang, Wang, and Chen (2013) described face representations using texture and geometrical shape. Previously, we also combined several different geometric features to represent ethnical groups, such as length, angular, and proportion features (Li et al., 2017). Also the semantic descriptions for ethnical groups were constructed based on Axiomatic Fuzzy Set (AFS) theory, and the manifolds of ethnical groups were learned in our recent study (Duan, Li, Wang, Zhang, & Liu, 2016; Wang, Duan, Liu, Wang, & Li, 2016).

SR has been intensively used in the field of face recognition, expression analysis, age estimation, and facial image super-resolution (Dian, Fang, & Li, 2017). Wright, Yang, Ganesh, Sastry, and Ma (2009) proposed sparse representation-based classification (ESRC) approach and brought the SR into face recognition. It assumed that a face image could be viewed as a sparse linear representation of other face images for the same person. Aharon, Elad, and Bruckstein (2006) applied the ESRC approach directly for occluded facial expression recognition. The performance is not as good as expected due to the fact that the identity information of human face is more obvious than that of expression, which implies that the features of identity would affect facial expression recognition severely. Recently, the SR has been extended to some recognition tasks with small sample size. Mairal, Leordeanu, Bach, Hebert, and Ponce (2008) proposed the extended ESRC approach, which has refined SR by adding general learning in the framework of ESRC. This method improved the performance for small sample size face recognition problem and single sample based face recognition problem by unitizing the information extracted from other datasets. Yang, Zhang, Yang and Zhang (2010) proposed the sparse variation dictionary learning (SVDL) approach, in which one could obtain the projection matrix according to a training set. The SVDL was then embedded in ESRC to conduct face recognition. However, SVDL needs plenty of training data which contained all type of images for each class to learn an effective dictionary. Yang, Zhang, Yang, and Niu (2007) proposed sparse illumination learning and transfer (SILT) approach. This approach could match a few targets for obtaining information of face images with different illuminations. The methods mentioned above can improve face recognition performance to different extent, and also have significant achievement in solving small sample size problem in face recognition. In this paper, we aim to use the SR approach to solve the ethnical group recognition with extracted regional features via data mining.

3 | THE MULTIETHNIC GROUP DATABASE

In order to investigate the ethnicity description and recognition, we collected a dataset including facial images of different ethnical students on campus in Dalian Minzu University, whose ages are ranging from 18 to 22 years old. The database includes three ethnicities, namely Korean, Tibetan, and Uyghur. The students of the three ethnicities are from the regions inhabited by the corresponding ethnical groups, as shown in Figure 2. For each ethnicity, 100 students are selected and their facial images are captured. The capture environment and setup are illustrated in Figure 3, in which we have three cameras, three lights with one person sitting in the center. The images of several participants are shown in Figure 4. Remember only the frontal images are used in this paper though we have collected images with different poses and expressions.

4 | FACIAL IMAGE PREPROCESSING

Due to the variations in pose, illumination, and camera parameters, it is necessary to align the images before further processing. This mainly involves face alignment and illumination normalization. The aim of face alignment is to correct face pose and resize the face resolution. The details are shown as follows:

- The coordinates of eyes are obtained automatically by eye detector, and the coordinates of two eye corners are denoted by E_l and E_r.
- The face image is rotated to make the line segment which connecting E_l and E_r to be horizontal.
- The facial area is cropped out according to the ratio of eye separation and the the rest of face.
- The cropped face images are resized to a given resolution.

As the skin color and texture of different ethnicities vary a lot, due to influence rendered by gene or environment, we conduct the illumination normalization in face image preprocessing stage. However, illumination normalization will affect the

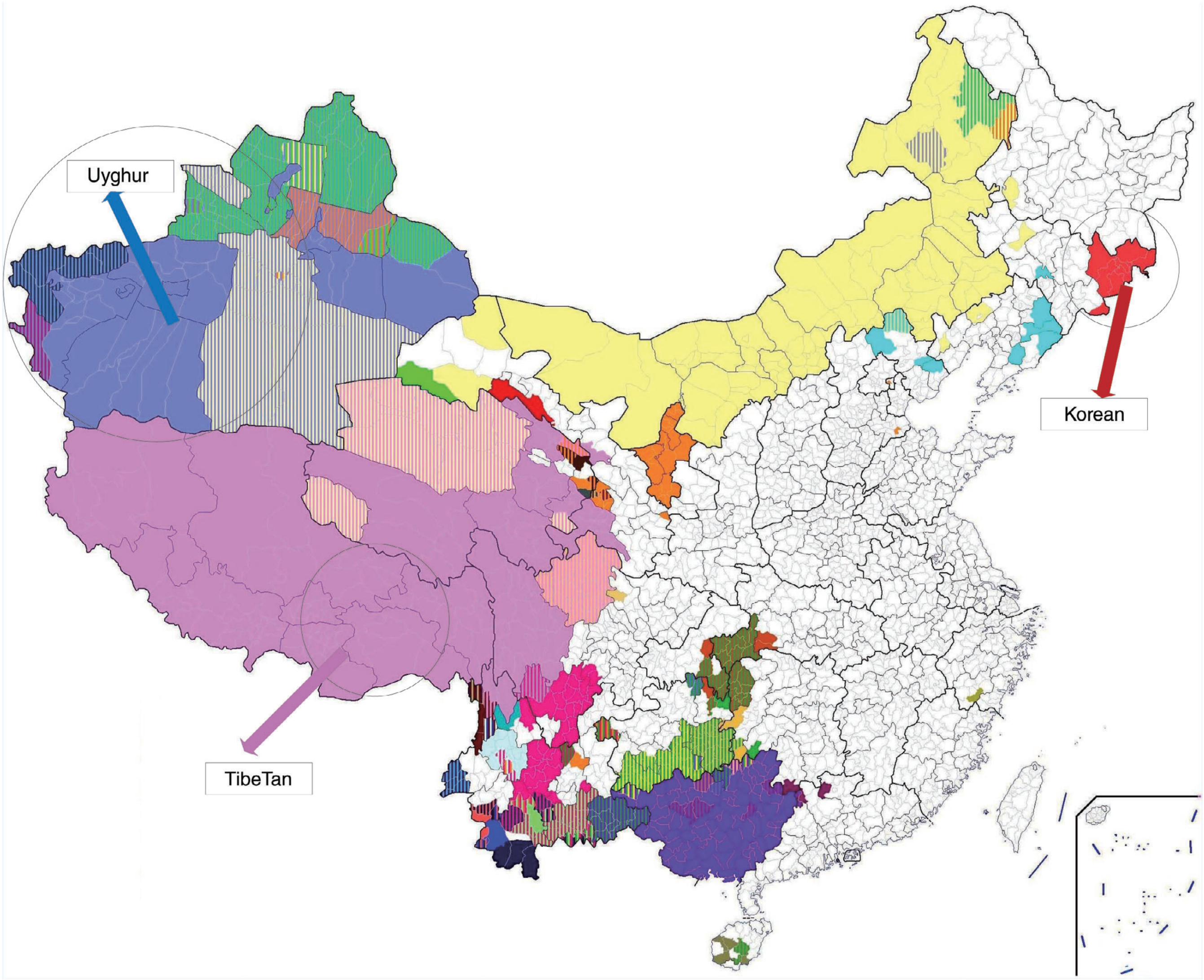

FIGURE 2 The living area distribution of three ethnicities

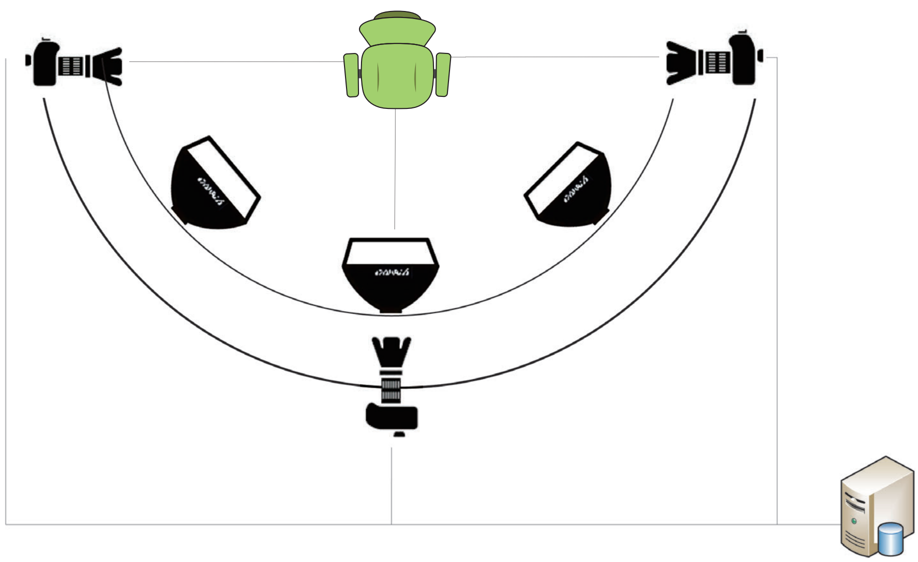

FIGURE 3 Data capture environment

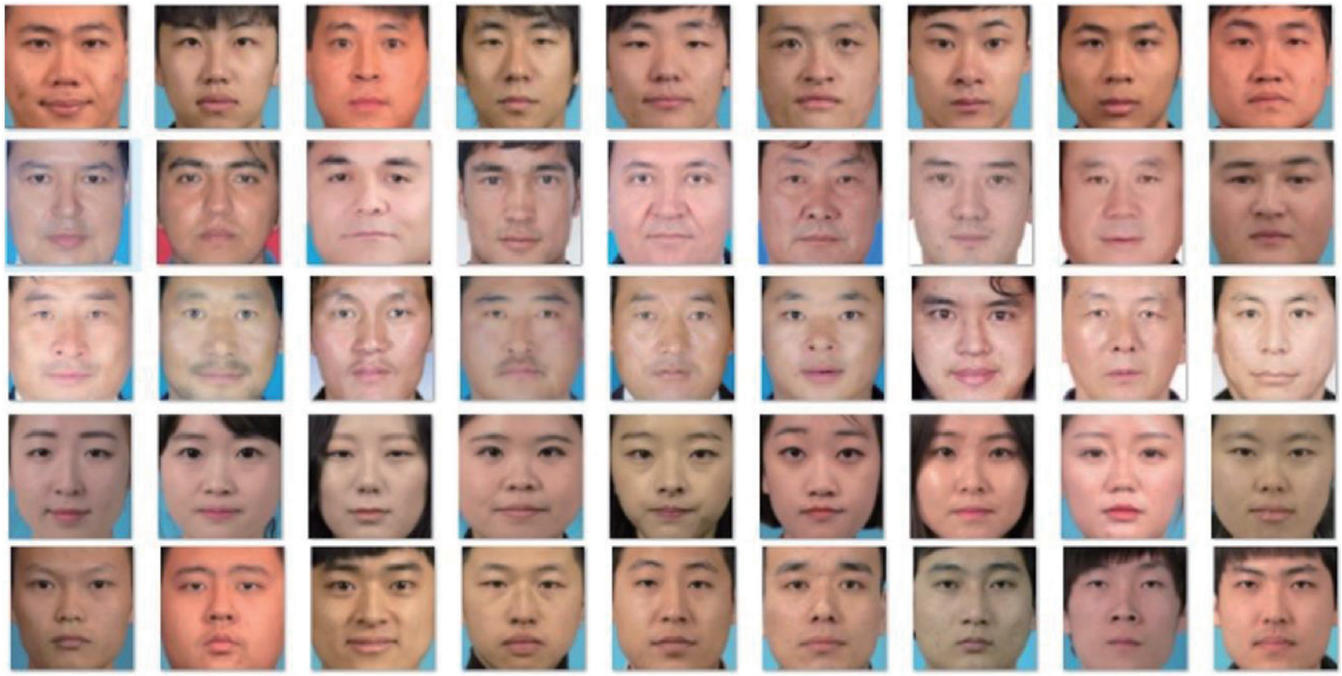

FIGURE 4 A part of face dataset

skin color. According to literature (Brooks & Gwinn, 2010), the skin color has a poor correlation with facial ethnic attributes. Hence, illumination normalization is implemented and the skin color change is ignored in our study.

Many methods have been proposed to deal with illumination variations (Biglari, Mirzaei, & Ebrahimpour-Komeh, 2013) such as single scale retinex (SSR), multiscale retinex (MSR), and homomorphic filtering (HOMO). In this paper, the SSR is used to normalize the illumination variations for simplicity. Suppose that the light is smoothly distributed over space, the brightness of object depends on the lighting of environment and reflection of objective surface, as shown in formula (1),

$$S(x,y) = R(x,y) \cdot L(x,y), \tag{1}$$

where $S(x, y)$ is the facial image captured by camera, $L(x, y)$ indicates component of lighting, and $R(x, y)$ represents reflection components of object. In order to separate reflection components and lighting components, logarithm operation is used as follows:

$$\log[R(x,y)] = \log \frac{S(x,y)}{L(x,y)}, \tag{2}$$

where $R(x, y)$ is corresponding to the high-frequency components of image, $L(x, y)$ represents the low-frequency components of image. In order to obtain $R(x, y)$, the Gaussian filter (Hyvrinen, Hoyer, & Oja, 1999) is then applied to estimate $L(x, y)$ as follows.

$$r(x,y) = \log[R(x,y)] = \log[S(x,y)] - \log[S(x,y) \times G(x,y)], \tag{3}$$

where $G(x, y)$ is the Gaussian function with $G(x,y) = K \times e^{-\left(x^2 + y^2/c^2\right)}$; c is the scale of Gaussian function; x is the size of Gaussian kernel; y is standard deviation of Gaussian distribution; K is a constant, and it satisfies $\int\int G(x, y)\mathrm{d}x\mathrm{d}y = 1$.

As shown in Figure 5, the experimental results demonstrate that SSR not only has good performance in illumination normalization, but also has a quick computational speed.

5 | THE ETHNICITY RECOGNITION USING SPARSE SENSING

5.1 | The kNN-based fast sparse sensing for ethnicity recognition

The SR for facial ethnicity recognition contains two steps. First, the K-nearest neighbors of a sample are selected from the whole training set for each group. Second, the sample is described and catergrized by the selected K-nearest neighbors via SR. The testing sample is described as a linear combination using its K-nearest neighbors (Waqas, Yi, & Zhang, 2013).

FIGURE 5 The results of face image using single scale retinex

The proposed fast SR algorithm consists of three steps: K-nearest neighbors identification, linear representation, and classification. In K-nearest neighbors identification, the K-nearest neighbors are identified and the corresponding labels are recorded. If a training sample belongs to jth ($j = 1, 2, \cdots, L$) class, j is taken as the label. Suppose $\{x_1, \cdots, x_K\}$ are the K-nearest neighbors of a testing sample y, their labels could form a new set $\boldsymbol{C} = \{c_1, c_2, \cdots, c_d\}$. The number of elements in this set is less than or equal to L or K. That is to say, $\boldsymbol{C}$ is a subset of $\{1, 2, \cdots, L\}$.

The testing sample y could be represented as a linear combination of the K-nearest neighbors

$$y = a_1 x_1 + \cdots + a_K x_K, \tag{4}$$

where $a_i (i = 1, 2, \cdots, K)$ are the coefficients. The formula (4) can be rewritten as follows:

$$y = \boldsymbol{XA}, \tag{5}$$

where $\boldsymbol{A} = [a_1, \cdots, a_K]^T$, $\boldsymbol{X} = [x_1, \cdots, x_K]$. Our aim is to solve the minimum error between $\boldsymbol{XA}$ and y, subject to that the norm of A must be minimum. This optimization problem can be described by a Lagrangian function,

$$\begin{aligned} L(\boldsymbol{A}) &= \|y - \boldsymbol{XA}\|^2 + \mu \|\boldsymbol{A}\|^2 \\ &= (y - \boldsymbol{XA})^T (y - \boldsymbol{XA}) + \mu \boldsymbol{A}^T \boldsymbol{A}' \end{aligned} \tag{6}$$

where μ is a positive constant. According to Lagrangian method, $\boldsymbol{A}$ should satisfy that $\frac{\partial L(\boldsymbol{A})}{\partial \boldsymbol{A}} = 0$. Therefore, the optimal solution could be obtained as follows:

$$\boldsymbol{A} = \left(\boldsymbol{X}^T \boldsymbol{X} + \mu \boldsymbol{I}\right)^{-1} \boldsymbol{X}^T y, \tag{7}$$

where $\boldsymbol{I}$ is an identity matrix.

The class label of a testing sample will be estimated according to its K-nearest neighbors's weight contribution in the SR (Wang et al., 2016). Specifically, in K-nearest neighbors of a testing sample, the subset $\{x_s, \cdots, x_t\}$ belongs to rth$(r \in C)$ class, the contribution of rth class is described as follows:

$$g_r = a_s x_s + \cdots + a_t x_t. \tag{8}$$

The error between g_r and the testing sample is given in formula (9).

$$e_r = \|y - g_r\|^2, r \in C. \tag{9}$$

The smaller of the value of $e_r = \|y - g_r\|^2$, implies the greater of influence on the r class. The testing sample y is then classified as the class which has the greatest contribution. In addition, if all K-nearest neighbors are not from rth class, r does not belong to C. Hence, the SR will not classify the sample y as the rth class. Two kinds of measurement similarity are usually used in the SR. One is Euclidean distance (Aharon et al., 2006), the other is cosine measure,

$$s(y, x_i) = \|y\| \cdot \|x_i\| \cos(\theta_i), \tag{10}$$

where $s(y, x_i)$ represents the similarity between y and ith neighbor.

5.2 | Holistic ethnical facial features based on SR

In this paper, the "O" region represents the whole face. We first implement SR based on holistic facial features, which implies that the whole image of a testing sample y is approximated by a linear combination of all the training images. The class label of a testing sample y is then assigned based on the difference between y and the weighted combination of the samples from each class. Let $\boldsymbol{A} = (\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n)$ denote n training samples, the testing sample y can be approximated as a linear combination of all training samples

$$y \approx \sum_{i=1}^{n} \boldsymbol{\beta}_i \boldsymbol{A}_i. \tag{11}$$

Without loss of generality, the formula is expressed as follows:

$$y = \boldsymbol{A}\boldsymbol{\beta}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)^T, \boldsymbol{A} = (\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n)$.

If $\boldsymbol{A}^T \boldsymbol{A}$ is nonsingular, the coefficients of $\boldsymbol{\beta}$ could be obtained by $\boldsymbol{\beta} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T y$. Otherwise, if $\boldsymbol{A}^T \boldsymbol{A}$ is singular, $\boldsymbol{\beta}$ could be calculated by $\boldsymbol{\beta} = (\boldsymbol{A}^T \boldsymbol{A} + \gamma \boldsymbol{I})^{-1} \boldsymbol{A}^T y$, where γ is a small positive number, and $\boldsymbol{I}$ is a unit matrix.

It can be seen from formula (11) that each training sample has its contribution to the representation of the testing sample, and the contribution of ith training sample is $\boldsymbol{\beta}_i \boldsymbol{A}_i$. Suppose the training samples from kth class are $\boldsymbol{A}_s, \cdots, \boldsymbol{A}_t$, and the total contribution of these samples to the testing sample y is denoted by $g_k = \boldsymbol{\beta}_s \boldsymbol{A}_s + \cdots + \boldsymbol{\beta}_t \boldsymbol{A}_t$, the error of the SR could be calculated by $e_k = \|y - g_k\|^2$. The smaller error value implies that the contribution is greater from the samples of kth class.

Now, we conduct a simple experiment for ethnicity recognition based on the holistic facial features on the captured dataset in this paper and the experimental results are shown in Table 1, the accuracy rate of ethnicity recognition is only 45% with 90% for training and 10% for testing on 10-fold experiments. It can be seen that the ethnicity recognition accuracy based on holistic facial features is quite low. In fact, the ethnic face is represented by sparse combination of various faces. However, one particular problem of ethnicity recognition is that the ethnic attributes come from various individuals but the facial attributes of individuals from different ethnicities may have significant contributions. In addition, holistic face features may contain insensitive features to ethnicity classification, since the facial ethnical differences are mainly conveyed by local features. Hence, it is important for us to figure out the local facial regions that are related to ethnic differences, and investigate whether the sparsity of such local features is useful for facial feature representation in terms of ethnicity recognition. We will investigate local feature extraction issue via data mining approach in next section.

5.3 | Salient ethnic facial region extraction

In this section, some salient ethnic facial regions will be investigated based on the three ethnicities. Since the geometric features are often used in anthropometry, this work also analyzes salient ethnic facial regions according to geometrical relationship of key points based on facial components. Here, we use the facial landmark detector STASM (Milborrow & Nicolls, 2014) to extract 77 landmarks as shown in Figure 6.

Based on these 77 landmarks, we can construct 2,926 facial features by connecting any two landmarks. Considering the redundancy and relevance of the obtained line features, the well-known data mining technique, the minimal-redundancy-maximal-relevance (mRMR) (Ding & Peng, 2005; Peng, Long, & Ding, 2005) feature selection method, is applied to select the most salient features. According to mutual information, the mRMR aims to select the significant features based on the minimal redundancy and maximal relevance, using the Equations (12) and (13),

$$\max D(F,c), D = \frac{1}{|F|} \sum_{f_i \in F} I(f_i, c), \tag{12}$$

$$\min R(F), R = \frac{1}{|F|^2} \sum_{f_i, f_j \in F} I\left(f_i, f_j\right), \tag{13}$$

where F is facial geometrical feature subset, c is class label of ethnicities, f_i is ith feature of F. $I(f_i, c)$ is mutual information between feature f_i and class c, and $I(f_i, f_j)$ indicates mutual information between f_i and f_j. The mutual information is calculated by Equation (14), and the mRMR selection criterion is achieved by the Equation (15).

$$I\left(f_i, f_j\right) = \int\int p\left(f_i, f_j\right) \log \frac{p\left(f_i, f_j\right)}{p(f_i) p\left(f_j\right)} \mathrm{d}f_i \mathrm{d}f_j, \tag{14}$$

TABLE 1 The ethnicity recognition based on holistic features

DataSet	TPR	FPR	Precision	Recall	*F*-measure	Accuracy
Holistic feature	0.39	0.61	0.24	0.39	0.35	0.45

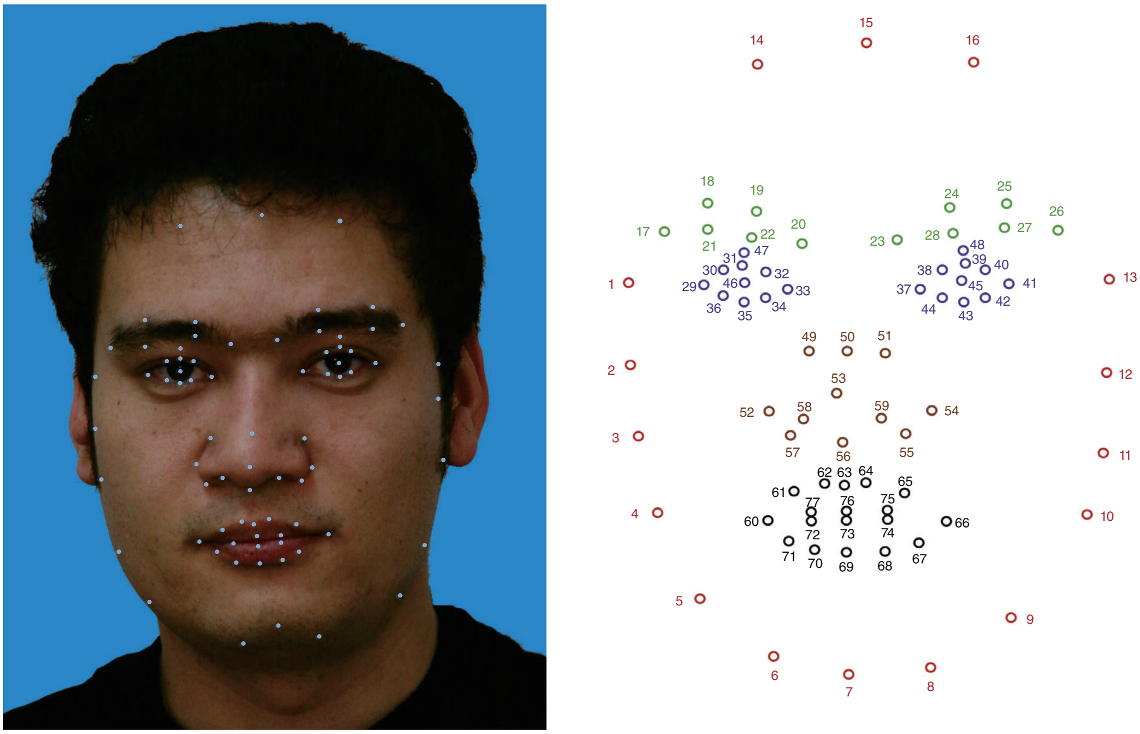

FIGURE 6 Landmarks obtained using STASM

$$\begin{cases} \max\varphi_1(D,R), \varphi_1 = D-R \\ \max\varphi_2(D,R), \varphi_2 = \dfrac{D}{R} \end{cases}. \tag{15}$$

Based on these 2,926 facial features, 195 salient length features are then selected out to represent the ethnic attributes of the three ethnicities using the mRMR approach. These features are divided into four parts, and then compared with anthropological features (Farkas, 1994). As shown in Figure 7, these four parts of the features are plotted on facial images. Figures 7a, b, c, and d show 19, 37, 63, and 65 length features, respectively. One can see that the best weights of features focus on nose, eyes, and eyebrows and these feature regions together form a "T" region, which can be seen clearly in Figure 7a and c. With the weights decreasing, the important region would extend to mouth area gradually. This observation shows that this "T" region is ethnic salient as demonstrated in next section.

5.4 | Local facial feature based ethnicity SR

From analysis in last section, the ethnic-salient "T" regions are first identified according to analysis of facial geometrical features, and the facial "T" regions are then used to recognize the ethnicity. As the shape of "T" regions are different in order to deal with various situations as described below, we propose three types of "T" regions, denoted as "T1," "T2," and "T3," which contain different facial components. As shown in Figure 8, "T1" includes eyes and nose, "T2" contains eyebrows, eyes, and nose, and "T3" contains eyebrows, eyes, nose, and mouth. Furthermore, the images of "T" regions are encoded according to zigzag rule for feature extraction in ethnicity recognition. "O" region represents the whole face image as explained below in Figure 9.

In the following analysis, the feature vector of "T" region is extracted to represent ethnic attributes. The K-nearest neighbors of a testing sample are selected based on the features from the corresponding "T" regions. The SR approach mentioned in previous section is then implemented to describe ethnicity attributes, which only locate in these "T" regions. The detailed algorithm can be described as follows:

- The "T" regions are identified based on landmarks obtained by STASM.
- The facial images are divided into training set $\boldsymbol{X} = [x_1, x_2, \cdots, x_m]$, and testing set $\boldsymbol{Y} = [y_1, y_2, \cdots, y_n]$, where x_i and y_i are the feature vectors extracted from the corresponding "T" regions.
- The K-nearest neighbors of each testing sample are selected, and the training labels are recorded.
- The testing sample $y \in \boldsymbol{Y}$ is represented by a linear combination

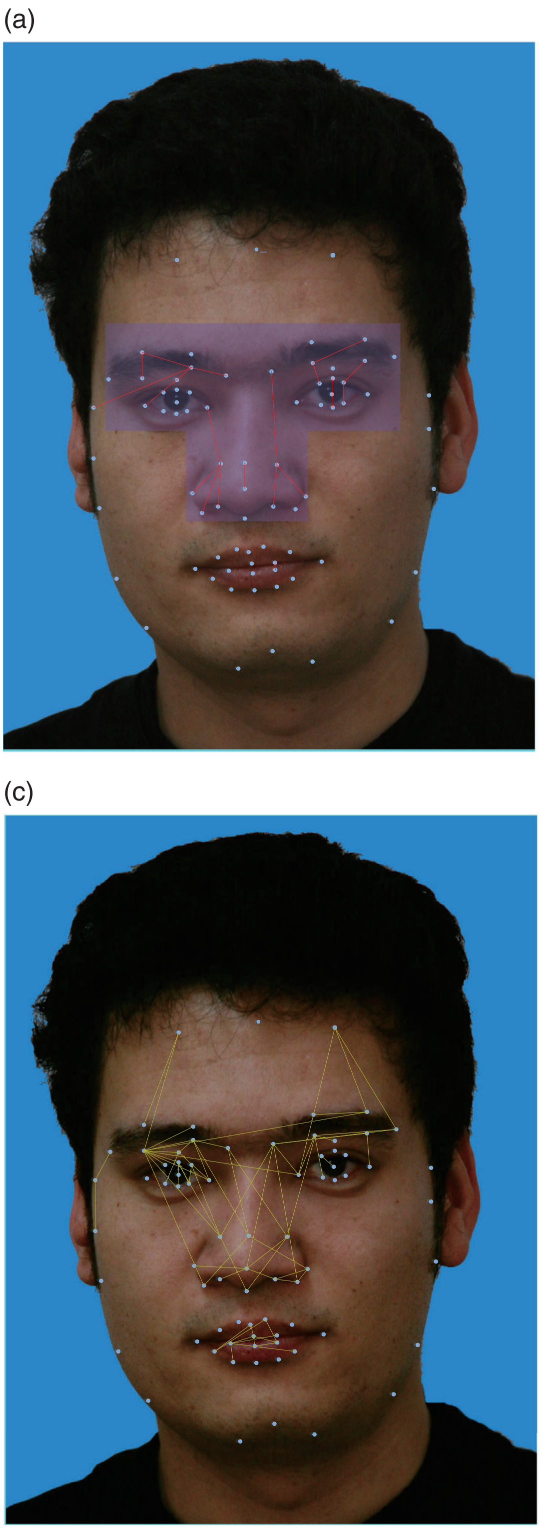

FIGURE 7 The various weight of length facial features

$$y = a_1 x_1 + \cdots + a_k x_k = \boldsymbol{X}\,\boldsymbol{A}. \tag{16}$$

- According to Lagrange optimization, the problem (16) could be solved, the optimal solution is given by:

$$\boldsymbol{A} = \left(\boldsymbol{X}^T \boldsymbol{X} + \mu \boldsymbol{I}\right)^{(-1)} \boldsymbol{X}^T y. \tag{17}$$

- The contribution of every class could be calculated:

FIGURE 8 Facial feature region of various weights

$$g_r = a_s x_s + \cdots + a_t x_t. \tag{18}$$

- The error between y and g_r could be obtained, and the class label of y is then identified according to the error e_r.

$$e_r = \|y - g_r\|^2, r \in C, \tag{19}$$

where we can select different norms in (19).

In summary, In order to represent each ethnical group effectively, we have used the STASM facial landmark detector to extract 77 landmarks in each facial image. Then, we construct 2,926 geometrical facial features. As the number of these features is too large, we used the data mining approach mRMR to select some salient geometrical features for these three ethnical groups and then 195 salient features are selected. One can find that these salient features are mainly located in a "T" region and then three types of "T" regions are constructed. We believe the features in these "T" regions are more important for ethnical group recognition. In next section, we demonstrate the effectiveness of the proposed framework.

6 | EXPERIMENTAL RESULTS

In this section, we conduct several experiments on the face images of Uygur, Tibetan, and Korean, and four types regions, that is, "0," "T1," "T2," and "T3," are established to extract ethnic salient features via using the data mining technique mRMR. The captured face images are first preprocessed, in which the faces are aligned and the illuminations are normalized. The effectiveness of the extracted features is then verified by using several different norms.

The performances of ethnicity recognition models on different "T" regions are evaluated by several different criteria, which include the true positive (TPR), false positive rate (FPR), Precision, Recall, and F-measure defined in (Anselmo, 1991; Bouckaert et al., 2010; Han, Pei, & Kamber, 2011). Next, we first conduct experiments on different "T" regions and validate the effectiveness of the proposed approach.

6.1 | The effectiveness of the three "T" regions

In this section, the coefficient number K is set to be 90, and the L_2 norm is applied in SR. Table 2 lists the results obtained based on different types of "T" regions. It can be seen that the result in the region of "T3" is the best among all regions. It

$$\begin{bmatrix} 26 & 3 & 6 & 2 & 2 & 1 & 3 & 23 & 3 & 13 \\ 0 & 3 & 4 & 1 & 1 & 6 & 31 & 12 & 13 & 4 \\ 3 & 1 & 5 & 12 & 31 & 2 & 4 & 4 & 14 & 1 \\ 4 & 1 & 19 & 1 & 1 & 1 & 32 & 2 & 2 & 2 \\ 1 & 3 & 2 & 2 & 8 & 2 & 1 & 11 & 1 & 3 \\ 5 & 11 & 1 & 8 & 13 & 3 & 8 & 2 & 2 & 1 \\ 4 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 11 & 2 \\ 6 & 4 & 3 & 1 & 6 & 2 & 2 & 2 & 2 & 1 \\ 2 & 1 & 5 & 2 & 3 & 6 & 3 & 1 & 12 & 3 \\ 3 & 4 & 6 & 1 & 4 & 3 & 2 & 2 & 2 & 4 \end{bmatrix}$$

FIGURE 9 The image coding of "T" region

TABLE 2 The results based on various "T" regions

Region	TPR	FPR	Precision	Recall	*F*-measure	Accuracy
O	0.390	0.610	0.240	0.390	0.350	0.450
T1	0.750	0.250	0.600	0.750	0.670	0.660
T2	0.780	0.220	0.630	0.780	0.700	0.690
T3	**0.865**	**0.135**	**0.762**	**0.865**	**0.810**	**0.780**

FPR, false positive rate; TPR, true positive rate.

indicates that the "T" region surrounded by eyes, eyebrows, nose, and mouth is more effective for ethnicity recognition than other "T" regions. Meanwhile, the results show that the region O is the worst in all regions, which is because the holistic facial images contain too much identity information rather than ethnic group features. Therefore, the SR based on local features is an effective approach to solve facial ethnic recognition. We believe the core contribution of "T" region via data mining plays an important role.

6.2 | Parameter selection

In order to study the influence of norms and *K*-neighbors, a series of different norms and *K*-neighbors are selected to identify ethnicities, and the recognition performances are compared based on the features from different facial regions. Specifically, the norms of L_0, L_1, and L_2 are adopted to evaluate recognition performance and the accuracy curves are plotted in Figure 10, Figure 11, and Figure 12, respectively.

Figure 10 shows the ethnicity recognition accuracy when the performance is evaluated by using L_0 norm. It can be seen that the best accuracy is achieved when the number of neighbors K equals to 77 and the features are extracted from "T3" region.

The results obtained based on the L_1 norm are shown in Figure 11. The recognition rate approaches to the peak based on the features from "T3" region when the neighbor number is 50. With the of neighbor number increasing from 15 to 50, the accuracy achieved based on "T3" region increases gradually with fluctuations. It suggests that "T3" region is the salient region for ethnic feature extraction and ethnicity recognition.

Figure 12 presents the recognition results using the L_2 norm. It can be seen that the best recognition performance is achieved based on the features extracted from "T3" region when neighbor number K is 80. Compared with L_0 and L_1 norm shown in Figure 10 and Figure 11, the highest accuracy is achieved by using L_2 norm, which reveals that L_2 norm is more appropriate than the other two types of norms in facial ethnicity recognition. In addition, the experimental results show that the performance obtained based on the features from "T3" region is better than that of "T1" and "T2," which means that identifying ethnic salient region can improve the recognition rate significantly.

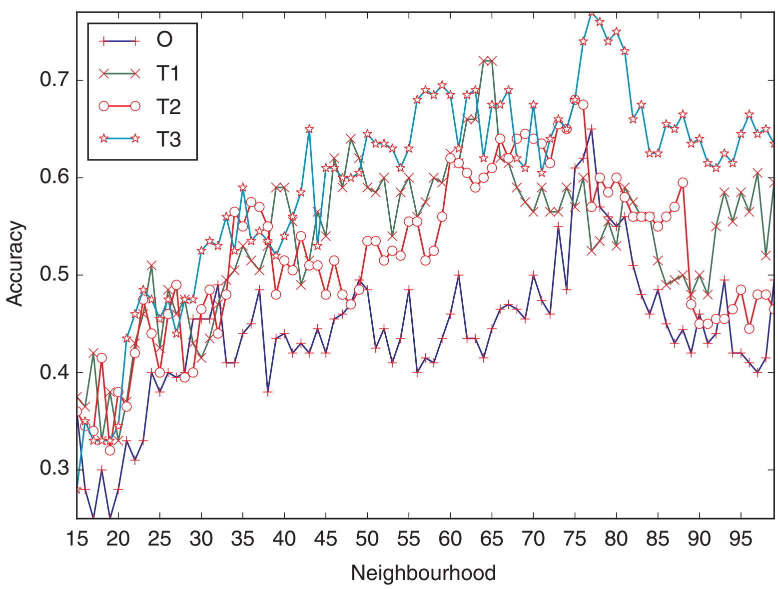

FIGURE 10 The accuracy based on L_0 norm

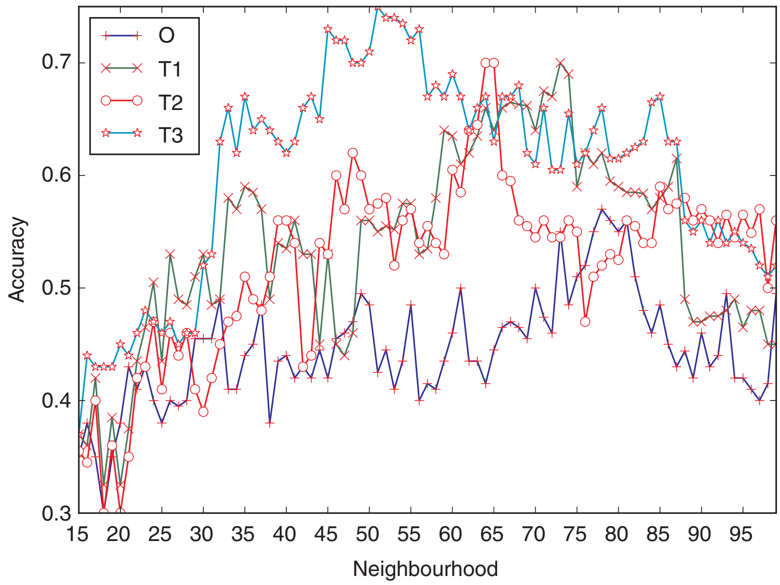

FIGURE 11 The accuracy based on L_1 norm

In summary, one can see that the proposed "T3" region is the most effective region for ethnicity recognition in combination with the L_2 norm using the SR. Next, we will develop a software platform for the visualization of ethnic facial feature description.

6.3 | Facial ethnic feature description

Based on previous analysis, this work attempts to describe the ethnic attributes according to the contribution of testing samples. As shown in Figure 13, a facial ethnicity evaluation system is constructed based on the SR coefficients. The k-nearest neighbors of a testing image on the left are determined based on the feature vector extracted from the "T" region. The SR coefficient (*coe*), the distance from the testing image to its k-nearest neighbors (*dis*), and the ethnicity identity of the testing image (*type*) are then obtained accordingly.

The error distance (*err*) from the testing sample to its k-nearest neighbors could serve as an important reference for facial ethnic description. As illustrated in Figure 14, the error distance *err* of a testing sample to the ethnic of Uyghur male is 0.01992, which means the most possible ethnic category of this sample is Uyghur. It can be seen from Figure 13, the k-nearest neighbors of this testing sample belong to several different ethnicities, its ethnicity could be estimated more precisely based

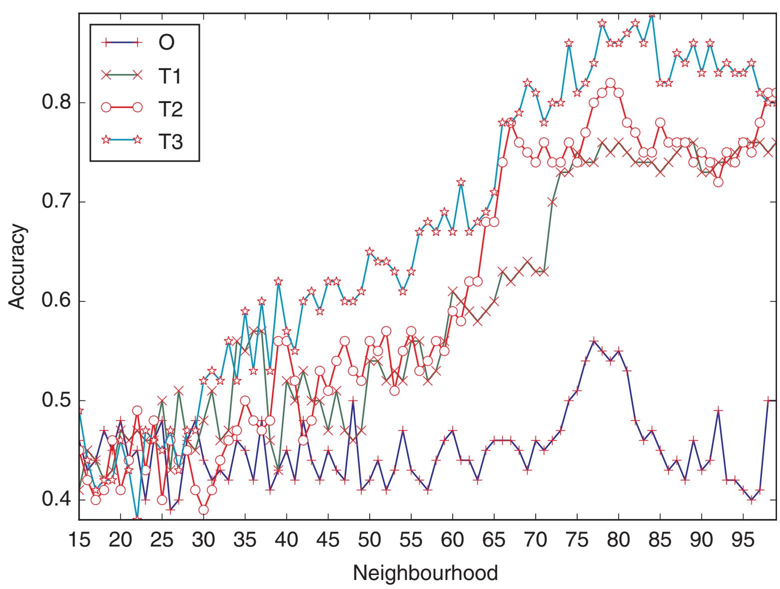

FIGURE 12 The accuracy based on L_2 norm

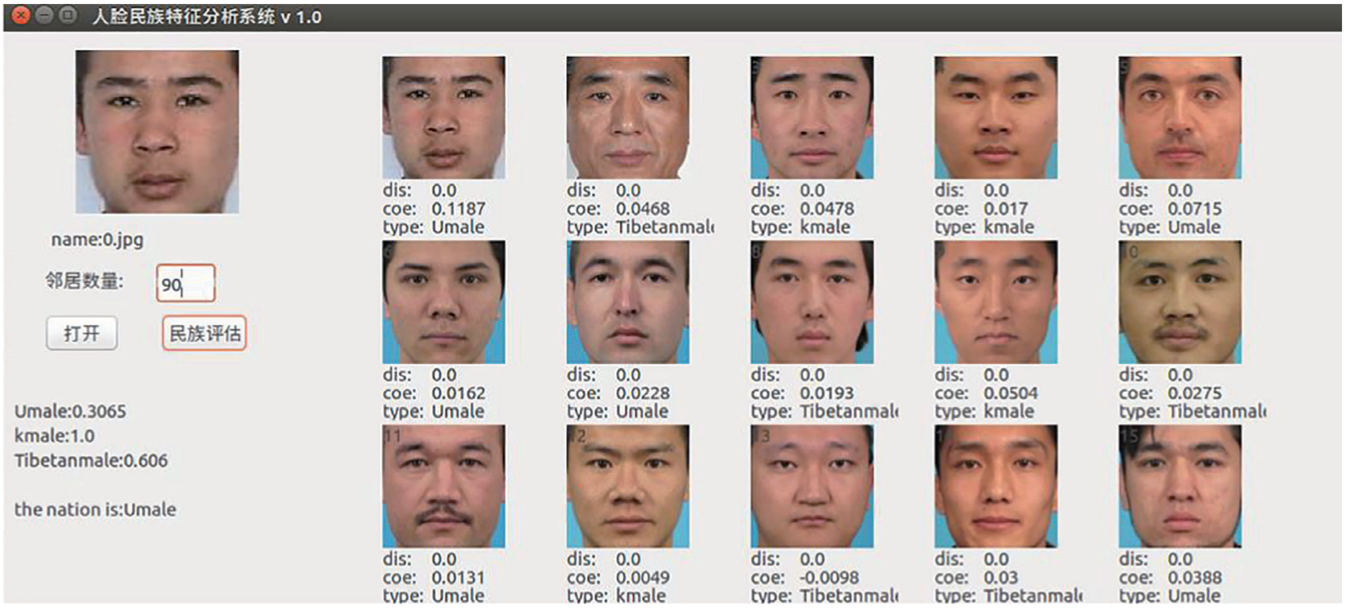

FIGURE 13 The software for face ethnic analysis

on the error distance *err*. Therefore, the ethnicity recognition depends on ethnic features in the constructed "T" region. It should be reminded that the error distance in the software platform is normalized for easy use.

6.4 | The investigation of the "T" region for face recognition

In previous section, one can see that the facial "T" region has shown its effectiveness in ethnicity recognition. Thus, it is straightforward to ask whether it is also useful for face recognition. In order to answer this question, some experiments of face recognition are conducted on olivetti research laboratory (ORL) database Figure 15 (Samaria, Harter & Harter, 1994). ORL database includes 400 face images of 40 persons with minor pose variations, and has been used for face recognition algorithm evaluation for decades. Since it is lack of pattern variation, the recognition rates for many face recognition systems have exceeded 90%.

The facial images of ORL database are divided into training and testing set, and the feature vectors are extracted from "T3" and "O" regions separately. The fast sparse classification based on k-nearest neighbors is also used to perform face recognition, and the results are shown in Figure 16 and Table 3. It can be seen clearly that recognition rate obtained based on holistic face ("O" region) is much better than that obtained based on local region ("T3" region). When k = 90, the recognition

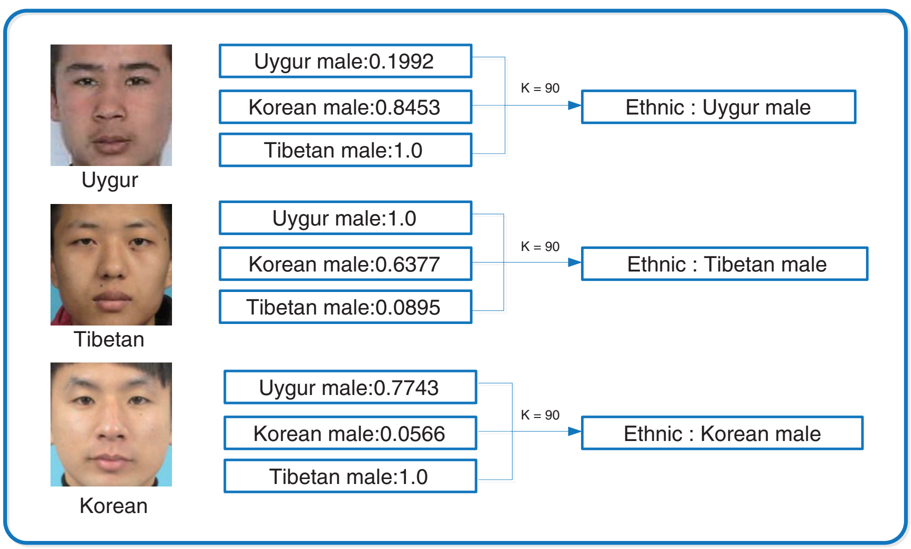

FIGURE 14 The results of classifiers

FIGURE 15 Olivetti research laboratory face dataset

rate based on "T3" region is only 63% but the accuracy based on holistic face has reached 90%. In fact, the performance achieved based on "T3" region never exceeds 70%, no matter how many neighbors are taken into consideration.

The experimental results indicate that the constructed "T" region is only suitable for ethnicity identification and not suitable for face recognition. This is mainly caused by the differences in samples referred in the SR. The referred samples of ethnicity are consisting of different individuals, while the referred samples of face recognition are from one individual with different poses and expressions. Moreover, the ethnic salient information concentrates in the "T" regions, but the information

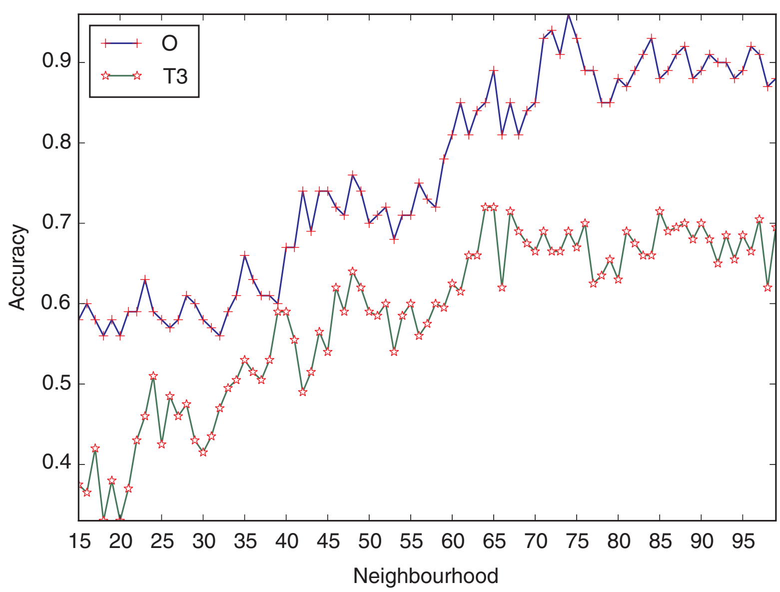

FIGURE 16 Use of the olivetti research laboratory database for testing

TABLE 3 Different T-zone recognition results for olivetti research laboratory datasets

Region	TPR	FPR	Precision	Recall	*F*-measure	Accuracy
O	0.94	0.06	0.81	0.94	0.87	0.90
T3	0.70	0.20	0.55	0.70	0.53	0.63

FPR, false positive rate; TPR, true positive rate.

enclosed in "T" regions is not enough for general face recognition. Actually, the facial features extracted from the "T" regions are more suitable for ethnicity recognition since the unrelated information has been filtered out.

7 | CONCLUSIONS

This paper aims to extract salient features via data mining for ethnicity recognition. First, the features extracted from holistic facial images are utilized for ethnicity recognition, and the recognition rate is quite low. This is because the facial ethnic features are different from the features extracted for face recognition. Consequently, this work continues to extract salient regions for ethnicity recognition. For such purpose, this work detects 77 facial landmarks to construct features for ethnicity representation according to anthropometry. The distance between each pair of landmarks is used to form a feature set, and 2,926 length features are produced for ethnical group description and then 199 features are selected after mRMR feature selection. Second, based on the selected features via using the data mining technique mRMR, three "T" regions including the most salient ethnic features are constructed. The experiments are conducted based on the features extracted from holistic face, "T1," "T2," and "T3" regions, the results show that the features from "T3" region would achieve the best performance when L_2 norm is adopted. Third, in order to verify the suitability of "T" region in face recognition, the facial features are extracted from "T" region on ORL dataset, and the fast sparse classification approach based on *k*-nearest neighbors is used to conduct face recognition, and the results suggest that the proposed "T" region is not suitable for face recognition.

The contributions of this paper are as follows: (a) The holistic facial features are proved to be ineffective for ethnicity analysis and recognition based on sparse sensing recognition. (b) The ethnic salient "T" region is proposed for ethnic attribute description via data mining technique. (c) The effectiveness of "T" region to ethnicity classification is verified. (d) The application of "T" region is investigated, it is suitable for ethnicity recognition but not for face recognition. In addition, this paper proposes a new approach for extracting facial ethnic features based on sparse description via data mining. The testing samples are sparsely represented and then assigned with its ethnic category accurately even under small sample size circumstance. Meanwhile, a framework for facial features analysis is proposed, that is, a framework for salient area search based on the data-driven feature selection, which can improve the effectiveness of the attribute discrimination using SR.

In the future, we will use different approaches instead of the SR to investigate this ethnicity recognition problem, which is different from general face recognition as shown in this paper. One possible direction is to extract the geometric features in the identified "T" region and use some deep learning (Pathirage, Li, & Liu, 2017) or stochastic configuration neural networks for classification (Wang & Li, 2017).

ACKNOWLEDGMENTS

We thank the two reviewers and Associate editor for their constructive comments and the quality of this paper is significantly improved after careful revision based on their comments. At same time, this work is supported by the National Natural Science Foundations of China with grant number 61562093 & 61772575 and the China Education & Research Network Innovation projects with grant numbers NGII20170419 & NGII20170631.

RELATED WIREs ARTICLES

Multidimensional compressed sensing and their applications
Sparse Models for Computer Vision

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

REFERENCES

Aharon, M., Elad, M., & Bruckstein, A. (2006). *rmk*-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, *54*(11), 4311–4322.

Akbari, R., & Mozaffari, S. (Mar. 2012). Fuzzy moments method for face recognition in an ethnic database. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *5*(1), 85–98.

Anselmo, M. (1991). Automates et codes zigzag. Automata and zigzag codes. (automates et codes zigzag.). *RAIRO—Theoretical Informatics and Applications*, *25*(1), 49–66.

Anzures, G., Pascalis, O., Quinn, P. C., Slater, A. M., & Lee, K. (2011). Minimizing skin color differences does not eliminate the own-race recognition advantage in infants. *Infancy*, *16*(6), 640–654.

Awwad, A., Ahmad, A., & Salameh, W. A. (2013). Arabic race classification of face images. *International Journal of Computers & Technology*, *4*(2), 397–402.

Biglari, M., Mirzaei, F., & Ebrahimpour-Komeh, H. (2013). Illumination invariant face recognition using SQI and weighted LBP histogram. *Pattern Recognition and Image Analysis (PRIA)* (Vol. 7, pp. 1–7), IEEE.

Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2010). Wekaâ ˇexperiences with a java open-source project. *Journal of Machine Learning Research*, *11*, 2533–2541.

Brooks, K. R., & Gwinn, O. S. (2010). No role for lightness in the perception of black and white? Simultaneous contrast affects perceived skin tone, but not perceived race. *Perception*, *39*(8), 1142–1145.

Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(10), 1042–1052.

Chan, H., & Bledsoe, W. (1965). *A man-machine facial recognition system: Some preliminary results*. Technical report, *In Panoramic research Inc*. Palo Alto, CA, USA1965.

Chen, S., Zhang, C., Dong, M., Le, J., & Rao, M. (2017). *Using ranking-cnn for age estimation.* Paper presented at IEEE Conference on Computer Vision and Pattern Recognition (pp. 742–751).

Cunrui, W., Qingling, Z., Xiaodong, D., Yuangang, W., & Zedong, L. (2018). Reserarch of face ethnic features from manifold structure. *Journal of Automation*, *44*(1), 140–159.

Dian, R., Fang, L., & Li, S. (2017). *Hyperspectral image super-resolution via non-local sparse tensor factorization.* Paper presented at IEEE Conference on Computer Vision and Pattern Recognition (pp. 3862–3871).

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, *3*(02), 185–205.

Ding, H., Huang, D., Wang, Y., & Chen, L. (2013). F*acial ethnicity classification based on boosted local texture and shape descriptions.* Paper presented at Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on IEEE (pp. 1–6).

Duan, X. D., Li, Z., Wang, C., Zhang, Q., & Liu, X. (2016). Multi-ethnic face eemantic description and mining method based on AFS. *Chinese Tournal of Computers*, *39*(7), 1435–1449.

Farkas, L. G. (1994). Anthropometry of the head and face. *Annals of Occupational Hygiene*, *52*(4), 773–782.

Fu, S., He, H., & Hou, Z.-G. (2014). Learning race from face: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(12), 2483–2509.

Fu, S. Y., Yang, G. S., & Hou, Z. G. (2011). *Spiking neural networks based cortex like mechanism: A case study for facial expression recognition.* Paper presented at International Joint Conference on Neural Networks (pp. 1637–1642).

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. San Francisco: Margan Kaufmann Publishers.

Hyvrinen, A., Hoyer, P., & Oja, E (1999). Denoising of nongaussian data by independent component analysis and sparse coding. In *ICA99, IEEE Workshop on Independent Component Analysis and Blind Source Separation, Aussois.* CiteSeerX.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, *13*(12), 524–531.

Jianwen, H., Lihua, W., Lilongguang, L., & Shourong, C. (2010). Analysis of morphous characteristics of facial reconstruction and the five organs in chinese north five national minorities crowd. *Journal of Chongqing Medical University*, *35*(2), 297–303.

Kanade, T. (1977). Computer recognition of human faces. *Interdisciplinary Systems Research*. Basel, Switzerland: Birkhauser Verlag.

Kumar, N., Berg, A., Belhumeur, P. N., & Nayar, S. (2011). Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(10), 1962–1977.

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1364.

Li, Z., Duan, X., Zhang, Q., Wang, C., Wang, Y., & Liu, W. (2017). Multi-ethnic facial features extraction based on axiomatic fuzzy set theory. *Neurocomputing*, *242*, 161–177.

Luan, S., Zhang, B., Zhou, S., Chen, C., Han, J., & Liu, J. (2018). Gabor convolutional networks. *IEEE Transactions on Image Processing*, *27*(9), 4357–4366.

Mairal, J., Leordeanu, M., Bach, F., Hebert, M., & Ponce, J. (2008). Discriminative sparse image models for class-specific edge detection and image interpretation. In *European Conference on Computer Vision* (pp. 43–56). Berlin, Heidelberg: Springer.

Milborrow, S., & Nicolls, F. (2014). *Active shape models with sift descriptors and mars.* Paper presented at Computer Vision Theory and Applications (VISAPP), 2014 International Conference on IEEE (Vol. 2, pp. 380–387).

Ortiz, E. G., Wright, A., & Shah, M. (2013). Face recognition in movie trailers via mean sequence sparse representation-based classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3531–3538.

Pathirage, C. S. N., Li, L., & Liu, W. (2017). *Discriminant auto encoders for face recognition with expression and pose variations*," Paper presented at International Conference on Pattern Recognition (pp. 3512–3517).

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238.

Ptucha, R., Tsagkatakis, G., & Savakis, A. (2011). Manifold based sparse representation for robust expression recognition without neutral subtraction. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 2136–2143), Barcelona, Spain: IEEE.

Salah, S. H., Du, H., & Al-Jawad, N. (2013). Fusing local binary patterns with wavelet features for ethnicity identification. In *Proceedings of World Academy of Science, Engineering and Technology*, no. 79. World Academy of Science, Engineering and Technology (WASET), p. 471.

Samaria, F. S., & Harter, A. C., Harter, A. (1994). Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision* (pp. 138–142).

Shiyuan, H. (2002). Early meanings and application of the terms "ethnos" and "ethnic group". *Ethno-National Studies*, *4*, 1–10.

Srinivas, N., Atwal, H., Rose, D. C., Mahalingam, G., Ricanek, K., & Bolme, D. S. (2017). *Age, gender, and fine-grained ethnicity prediction using convolutional neural networks for the East Asian face dataset.* Paper presented at IEEE International Conference on Automatic Face and Gesture Recognition (pp. 953–960).

Sun, Y., Wang, X., & Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2892–2900), Boston, MA.

Trigeorgis, G., Snape, P., Kokkinos, I., & Zafeiriou, S. (2017). *Face normals "in-the-wild" using fully convolutional networks*. Paper presented at Computer Vision and Pattern Recognition (pp. 340–349).

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Wade, P. (2007). *Race, ethnicity, and nation: Perspectives from kinship and genetics* (Vol. 1). New York City: Berghahn Books.

Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., & Ma, Y. (2012). Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(2), 372–386.

Wang, D., & Li, M. (2017). Stochastic configuration networks: Fundamentals and algorithms. *IEEE Transactions on Cybernetics*, *47*(10), 3466–3479.

Wang, Y., Duan, X., Liu, X., Wang, C., & Li, Z. (2016). Semantic description method for face features of larger chinese ethnic groups based on improved wm method. *Neurocomputing*, *175*(PA), 515–528.

Waqas, J., Yi, Z., & Zhang, L. (2013). Collaborative neighbor representation based classification using l2-minimization approach. *Pattern Recognition Letters*, *34*(2), 201–208.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(2), 210–227.

Yan, Y., & Zhang, Y.-J. (2009). *State-of-the-art on video-based face recognition*. Paper presented at Encyclopedia of Artificial Intelligence (pp. 1455–1461). IGI Global.

Yang, J., Zhang, D., Yang, J.-y., & Niu, B. (2007). Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(4), 650–664.

Yang, M., Zhang, L., Yang, J., & Zhang, D. (2010). *Metaface learning for sparse representation based face recognition*. Paper presented at Image Processing (ICIP), 2010 17th IEEE International Conference on IEEE (pp. 1601–1604).

Zhang, Z., Song, Y., & Qi, H. (2017). *Age progression/regression by conditional adversarial autoencoder*. Paper presented at Computer Vision and Pattern Recognition (pp. 4352–4360).

How to cite this article: Wang C, Zhang Q, Liu W, Liu Y, Miao L. Facial feature discovery for ethnicity recognition. *WIREs Data Mining Knowl Discov*. 2019;9:e1278. https://doi.org/10.1002/widm.1278