Local circular patterns for multi-modal facial gender and ethnicity classification

Di Huang, Huaxiong Ding, Chen Wang, Yunhong Wang, Guangpeng Zhang, Liming Chen

Local circular patterns for multi-modal facial gender and ethnicity classification

Di Huang [a,*], Huaxiong Ding [b], Chen Wang [b], Yunhong Wang [a], Guangpeng Zhang [a], Liming Chen [b]

[a] State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, 100191 Beijing, China
[b] Department of Mathematics and Computer Science, Ecole Centrale de Lyon, CNRS, 69134 Lyon, France

Gender and ethnicity are both key demographic attributes of human beings and they play a very fundamental and important role in automatic machine based face analysis, therefore, there has been increasing attention for face based gender and ethnicity classification in recent years. In this paper, we present an effective and efficient approach on this issue by combining both boosted local texture and shape features extracted from 3D face models, in contrast to the existing ones that only depend on either 2D texture or 3D shape of faces. In order to comprehensively represent the difference between different genders or ethnicities, we propose a novel local de-scriptor, namely local circular patterns (LCP). LCP improves the widely utilized local binary patterns (LBP) and its variants by replacing the binary quantization with a clustering based one, resulting in higher discriminative power as well as better robustness to noise. Meanwhile the following Adaboost based feature selection finds the most discriminative gender- and race-related features and assigns them with different weights to highlight their importance in classification, which not only further raises the performance but reduces the time and mem-ory cost as well. Experimental results achieved on the FRGC v2.0 and BU-3DFE datasets clearly demonstrate the advantages of the proposed method.

1. Introduction

Gender and ethnicity[1] clues are two of the most fundamental and important demographic attributes of human beings, which remain unchanged all through lifetime. While people can easily recognize the gender and ethnicity of each other through their facial appearances, it is still a non-trivial problem for computers. Automatic face-based gender and ethnicity classification has promising applications in Human Computer Interaction (HCI), Business Intelligence (BI), surveillance, video and image retrieval, database indexing, and can provide useful information for face recognition.

1.1. Related work

During the last decade, research on face-based gender and ethnicity classification has grown up rapidly since it emerged. Most of the proposed techniques are 2D texture image based, as people from different genders and ethnicities commonly have a diversity of face textures. These methods can be approximately divided into three categories: i.e. raw image based, texture feature based, and geometry feature based.

Raw image based approaches consider the entire raw face image (generally frontal) as the input and employ dimensionality reduction techniques (down-sampling or subspaces) to process the facial image to finally feed the classifier. For instance, in the early 1990s, Golomb et al. [1] trained two neural networks, one for image compression, and the other for gender classification. Facial images were first normalized to the size of 30×30, and compressed to the dimension of 40, and then classified by the back-propagation SexNet. Their experiments on a set of 90 images (45 males and 45 females) displayed an average error rate of 8.1% compared to the one of 11.6% from the Psychophysical studies of five humans. Moghaddam and Yang [2] introduced a gender classification method, in which facial images were firstly down sampled into the size of 21×12, and then classified by Support Vector Machine (SVM). An error rate of 3.38% was achieved on a database containing 1755 samples. Lu and Jain [3] proposed an ethnicity classification method, in which facial images were analyzed at multiple scales, an LDA classifier was constructed for each scale, and the final matching score was obtained by fusing all similarity measurements under product rule. A classification accuracy of 96.3% was achieved on a union database containing 2630 samples of 263 subjects.

Texture feature based approaches describe local texture changes of certain areas to discriminate male from female or distinguish different ethnicities. Due to its low computational complexity, Shakhnarovich

* Corresponding author. Tel.: +86 10 82338431.
E-mail address: dhuang@buaa.edu.cn (D. Huang).

[1] We use "race" and "ethnicity" interchangeably here as the same concept is expressed by both terms in the previous literature.

et al. [4] investigated Haar-like features in combination with Adaboost for gender and race classification. Later, Hosoi et al. [5,6] extracted Gabor features and adopted SVM for the same issue. Approximately an accuracy of 91.6% was obtained for gender classification on a database with 1240 facial images, while the one of 94% was achieved for ethnicity classification using 1991 face photos. Lian et al. [7] improved SVM by the proposed Min–Max Modular SVM (M^3-SVM) to classify Gabor features, and M^3-SVM performed better (around 6% higher) than SVM in gender classification on 12,912 probes with the variations in facial expression, pose, occlusion, etc. In 2007, Lu and Lin [8] compared boosted Haar-like and Gabor features with SVM to recognize sex on 518 frontal facial images of size 24×24 from the FERET database, and pointed out that Gabor features were more effective than Haar features for the given task. In the same year, Yang and Ai [9] proposed an LBP (local binary patterns) based method for demographic attribute classification, and they showed that with the help of AdaBoost, LBP features achieved promising results in both tasks of gender and ethnicity classification on the subsets chosen from FERET and PIE, prior to the ones reached by Haar features. Guo and Mu [10] used Gabor features and reported benchmark performance for classification of five ethnicities on large-scale dataset MORPH-II containing more than 55,000 facial images. The ethnicity prediction for the Black and White races is 98.3% and 97.1%, while because of insufficient training data, for the other three races: Hispanic, Asian and Indian, the predictions are degraded dramatically to 74.2%, 59.5%, and 6.9%, respectively.

Geometry feature based approaches measure shape variations (including angles, distances and areas) of different genders and ethnicities respectively based on the 2D spatial arrangement of a set of facial fiducial points, such as the nosetip, the inner and outer corners of eyes, the endpoints of eyebrows, and claim that the clue conveyed in face shape also largely contributes to classify gender as well as race. Brunelli and Poggio [11] selected sixteen geometry features and used them in gender classification, a correct classification rate of 79% was reached. Samal et al. [12] computed 406 geometry features and noted that around 85% of features show significant difference between male and female.

Up till now, most efforts have been made within the 2D domain, i.e. using texture information. However, according to the anatomical studies, 3D geometrical information of faces of human beings also reflects distinctions among races and genders, and is thereby essential for gender and ethnicity classification as well. For example, faces of white people and male are commonly craggier than that of Asian people and female. Caucasian brow bones are always deeper, with eyes more sunken than Asian ones; while Asian noses tend to possess lower bridges; Caucasian noses extend slightly upward. Due to the development of 3D imaging technologies, 3D shape information of human faces can be easily captured, which facilitates the advance in 3D shape-based approaches. O'Tool et al. [13] applied Principal Component Analysis (PCA) to 3D coordinates of 130 faces to extract features, and reported a peak correct rate of 96.9% in a 17-dimensional subspace, better than the best score of 93.8% achieved based on their corresponding gray-level images in a 20-dimensional subspace. They further pointed out that by combining both modalities at the feature level, final performance was improved to 97.7% with a minimum subspace of 32 coefficients (half 3D based and half 2D based). Han et al. [14] manually selected some regions on 3D face meshes, and calculated the ratio of surface area and volume of these regions in comparison to the whole face as features for gender classification. An error rate of 17.44% was obtained by an SVM classifier on the GavabDB database. Wu et al. [15] proposed a supervised method namely weighted Principal Geodesic Analysis (PGA) to extract gender discriminating features from 2.5D facial needle-maps, and an accuracy of 97% was obtained on the Max–Planck face database which comprises 200 facial sans. Hu et al. [16] separated 3D faces into five regions, and introduced SVM to gender classification. The final result was obtained by combining the similarity scores of these five regions. An accuracy of 94.3% was achieved on a mixed database with 945 face models. Toderici et al. [17] adopted a high-level demographic feature estimated from the 3D meshes of the human face, and an accuracy of 99.6% was achieved for a two-class (Asian vs. White) classification on FRGC v2.0.

1.2. Motivation and contribution

Since most of the current 3D imaging systems deliver 3D face models along with their aligned texture counterparts, a major trend in the literature of face recognition is to adopt both the 3D shape and 2D texture based modalities, arguing that the joint use of these two clues can generally achieve more accurate and robust accuracy than using only either of the single modality [18]. We believe that fusion of 2D and 3D data will improve the classification accuracy in the classification of gender and ethnicity; nevertheless, very limited research has investigated this topic using multiple-modalities. Lu et al. [19] can be regarded as the pioneer for this attempt where they integrated similarity measurements of texture and shape (i.e. intensity and depth value of the central part of human faces), showing that the combination of multiple modalities leads to an improvement in both the accuracies of gender and race classification. This work is somewhat intuitive, and it has several downsides. For example, the direct use of pixel values of facial intensity and range images cannot sufficiently discriminate the difference between male and female or between various races, and it also tends to incur the sensitivity to illumination for the texture modality. In addition, they treat the entire face area equally, which is actually inappropriate.

Our basic assumption, as the one behind multi-modal face analysis, is that, the result of single modality (i.e. only 2D texture or 3D shape) based techniques can be ameliorated by combining various clues from different modalities. In this paper, we propose an effective and efficient approach to multi-modal face based gender and ethnicity classification, which addresses two important problems involving the process of extraction and selection of gender and ethnicity related features.

For the former, inspired by some recent studies on local feature based face recognition [20,21], in contrast to the original pixel values of facial texture and range images holistically used in [19], we investigate the way to represent the information of texture and shape in a local feature space with more discriminative power, aiming to minimize within-class variations and maximize between-class similarities. Due to its tolerance to monotonic lighting changes as well as computational simplicity, LBP is regarded as one of the most effective and successful local descriptors in many fields including texture analysis, facial image analysis, image and video retrieval, environment modeling, visual inspection, motion analysis, biomedical image analysis, aerial image analysis, and remote sensing [22]. However, LBP still has several main limitations, such as the insufficiency in discriminative power and the insensitivity to noise. In this work, rather than explicitly quantizing the sign and magnitude components of local patterns as adopted in LBP and its variants, we propose to quantize local patterns through clustering, and the descriptor is called local circular patterns (LCP for simplicity). Compared with the binary quantization scheme, clustering based quantization can generate better approximation with less distortion, therefore LCP possesses a greater ability in discrimination and is less sensitive against noise. Moreover, the quantization accuracy can be manageable through modifying the number of clusters. Additionally, in clustering based quantization, the parameters are tuned using training data, and it thus can deal with various local pattern distributions.

For the latter, among various feature selection techniques [23] proposed in the community of machine learning, we make use of the widespread Adaboost algorithm, to select a compact subset of facial features from the entire multi-modal feature set. The reason to employ Adaboost is that it is capable of obtaining a strong classifier through combining several weak ones while selecting features, as a result there is no need to retrain a classifier for gender or ethnicity label prediction. On the one hand, the features extracted from various facial regions (such as the ones of eyes, nose, forehead) represented as textures or shapes, highly related and discriminative to the task of facial gender and ethnicity classification, can be determined and assigned with different weights

according to their importance to final performance. The relevance of the features is thereby largely decreased, and the combination of these selected ones tends to improve the classification accuracy. On the other hand, the entire feature set generally contains redundant information, and utilizing all the features is also time and memory expensive which probably gives rise to the problem of curse of dimensionality. After feature selection, the dimensionality of the feature can be reduced and the efficiency of classification can be increased.

To sum up, this paper explores the discriminability of both 2D and 3D facial features in gender and ethnicity classification, and proposes multi-resolution local circular patterns (LCP) to represent local shape and texture variations. Histogramsof LCP features are extracted hierarchically, and Adaboost is used to select the most discriminative features from the high dimensional ones, while boosting weak classifiers into a strong one. Decision level fusion is made for final decision. Competitive performance is achieved on the FRGC v2.0 and BU-3DFE databases, highlighting the effectiveness of the proposed approach.

The remainder of the paper is organized as follows. Section 2 introduces LCP features in detail and highlights its improvements to LBP based ones. Section 3 presents the feature selection process as well as the decision-level fusion. Experiments and results are displayed and discussed in Section 4. Section 5 concludes the paper.

2. Local binary patterns vs. local circular patterns

Since local circular patterns (LCP) can be regarded as a variant of local binary patterns (LBP), we will first recall the basic concept of the original LBP descriptor, and then introduce the proposed LCP and LCP based facial representation.

2.1. The LBP methodology

A basic LBP operator simply thresholds a 3×3 neighborhood by the value of the central pixel, and the sign of thresholded neighboring values can form a binary number, which is then transformed into a decimal number. This decimal number is treated as the label of the central pixel (Fig. 1(a)). We call this quantization scheme as binary quantization in the following. The histogram of the labels within a region is often used as a texture descriptor.

Formally, given a pixel at (x_c, y_c), the derived LBP decimal value is:

$$LBP(x_c, y_c) = \sum_{n=0}^{8} s(i_n - i_c) 2^n \tag{1}$$

$$s(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases} \tag{2}$$

where n covers the eight neighbors of the central pixel, and i_c and i_n are the gray level values of the central pixel and its surrounding ones respectively.

The basic LBP was later extended to multi-resolution and "uniform" [24]. Multi-resolution denotes that LBP can operate on any radius R and any number of pixels P within the neighborhood, as shown in Fig. 1(b). The uniform pattern is defined as a local binary pattern which contains at most two transitions between 0 and 1. The extended LBP is notated as $LBP^{u2}_{P,R}$, indicating that the operator works in a (P, R) neighborhood, and employs only uniform patterns and labels all the remaining patterns with a single bin. The authors of [24] pointed out that uniform patterns are fundamental patterns providing the vast majority of all 3×3 patterns present in the observed textures.

However, LBP still has some limitations. One of the most critical limitations is that it only extracts the sign between neighboring pixels, while ignoring the magnitude, leading to the deficiency in discrimination. To better describe local micro patterns, various improvements have been explored [25–28]. In spite of the performance which is better than that of the LBP descriptor, these aforementioned variants are mostly artificially designed, and to comprehensively encode useful information, more bits are required which tends to incur a large increase in the memory and computational expense. Another limitation of LBP lies in that its binary coding scheme is very sensitive to noise, and once a single bit of the code alters, the resulting decimal number changes seriously. In order to improve the robustness to noise, several variants of LBP have been presented [29,30]. Most of them intrinsically inherit the binary coding scheme of the original LBP, and cannot completely solve this problem. Besides, the distribution of local patterns within images taken under different scenarios varies greatly, for example in scene images and face images, and using the same quantization parameters (e.g. the same threshold) is therefore unsatisfied.

2.2. Local circular patterns

According to the analysis on the properties of LBP as well as its variants, we find out that the two vital limitations above are mainly caused by the binary quantization scheme. In this study, rather than explicitly quantizing the sign or/and magnitude components of local patterns, we propose to make the quantization of local patterns through clustering, aiming to generate better approximation with less distortion and thus leading to improvement in discriminative power and robustness to noise.

Specifically, as illustrated in Fig. 1(b), for each pixel i_c whose gray value is t with its P neighboring pixels i_n, $n = \{1, 2, \ldots, P\}$ (gray values are denoted as $\{t_1, t_2, \ldots, t_P\}$) located on the circular neighborhood at the radius of R, the corresponding code p of this local circular pattern is defined as $p(LCP_{P,R}) = (t_1 - t, t_2 - t, \ldots, t_P - t)^T$. Given N training local circular patterns p_i, $i = 1, 2, \ldots, N$, the K-means clustering algorithm is performed to find a partition $C = \{c_1, c_2, \ldots, c_k\}$ by minimizing the following function:

$$J(C) = \sum_{i=1}^{k} \sum_{p_j \in c_i} D\left(p_j, \mu_i\right) \tag{3}$$

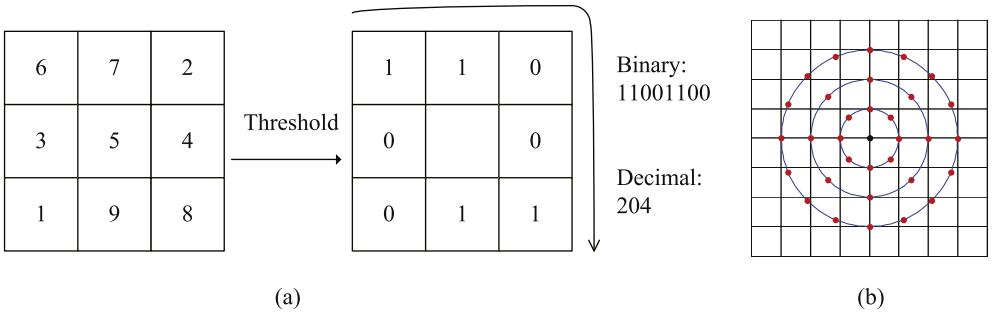

Fig. 1. LBP Operators: (a) Basic LBP; (b) Multi-resolution LBP.

where $D()$ represents the distance function, and μ_i is the center of c_i. Then a new local circular pattern p' can be quantized into the nearest cluster center.

$$l(p') = \arg\min_i D(p', \mu_i). \quad (4)$$

K-means is a greedy algorithm which can only converge to a local minimum, but the recent study has shown that K-means can converge to the global optimum with a large probability when clusters are well separated [31].

Two main issues associated with K-means clustering are the number of clusters as well as the distance metric. The number of clusters k controls the balance between descriptive power and sensitivity to noise. Reducing the number of clusters increases the distances among them, and little vibration of the pattern does not change its quantization, but the low number of clusters also reduces the descriptive power. K-means can be performed with various distances $D()$, and two of them are introduced in this study, namely the Euclidean distance (L2) and city block (L1) distance. Given a local circular pattern $p' = (p'_1, p'_2, \ldots, p'_P)$, the Euclidean distance between p' and a cluster center $\mu = (\mu_1, \mu_2, \ldots, \mu_P)$ is defined as:

$$D_{L2}(p', \mu) = \sqrt{\sum_{i=1}^{P} \left(p'_i - \mu_i\right)^2}, \quad (5)$$

and their city block distance is defined as:

$$D_{L1}(p', \mu) = \sum_{i=1}^{P} |p'_i - \mu_i|. \quad (6)$$

Euclidean distance is usually used for computing the distance between points and cluster centers. The clusters found by K-means with Euclidean distance are spherical or ball-shaped. K-means with city block distance was proposed in [32]. Compared with L2 distance, L1 distance is computationally more efficient. Each cluster center in the L1 distance case is calculated as the component-wise median of the points in that cluster. According to the clustering process, cluster centers obtained with L1 distance are all integers, while the ones obtained with Euclidean distance may be with decimals. In the following, in order to simplify the description, we call the LCP descriptor quantized by K-means with L2 and L1 distances LCP-L2 and LCP-L1 respectively.

2.3. LCP based facial representation

As we know, when LBP operates on the images formed by light reflection, i.e. 2D images, it can be used as a texture descriptor. Each of the LBP codes can be regarded as a micro-texton. Local primitives codified by the bins include different types of curved edges, spots, flat areas, etc. Meanwhile, as LBP works on range images which are based on depth information, it can also describe local shape structures [33], such as flat, concave and convex. Similar to LBP, to comprehensively represent facial texture and shape images, we follow the scheme proposed by Ahonen et al. [34] for 2D face recognition. The basic idea lies in that a face image can be considered as a composition of the micro-patterns described by an LBP-like descriptor. One can build an LBP-like histogram computed over the entire facial image. However, such a representation only encodes the occurrences of micro-patterns without any indication about their locations. In addition, to consider the configuration information of faces, face images can be divided into a certain number of local regions, from which local LBP-like histograms can be extracted. These histograms are then concatenated into a single, spatially enhanced feature vector. The resulting histogram encodes both the local texture and global shape of face images. In our case, as shown in Fig. 2, both 2D texture and 3D range images are aligned based on eye outer corners, and cropped by an average mask. They are then divided into m (to be fixed experimentally) rectangular regions. For each region, histograms of clustering quantization based LCP are extracted which are further concatenated into a single histogram as gender and ethnicity related features in both the texture and shape modalities.

2.4. Multi-scale extension

Some LBP histogram-based applications change the neighborhood of the LBP operator for improved performance. By varying the value of radius R, the LBP codes of different resolutions are obtained. The multi-scale strategy was originally used for texture classification [24], and it was also introduced to 2D face recognition [35,36]. In [37], Shan and Gritti studied MS-LBP for facial expression recognition by firstly extracting MS-LBP histogram-based facial features and then using the AdaBoost algorithm to learn the most discriminative bins. They reported that the boosted classifiers of MS-LBP consistently outperform those based on single-scale LBP, and the selected LBP bins distribute at all scales. MS-LBP can thus be regarded as an efficient method for facial representation. When considering it in multi-modal facial gender and ethnicity classification, this multi-scale technique can be applied to enhance the descriptive power of LCP as well.

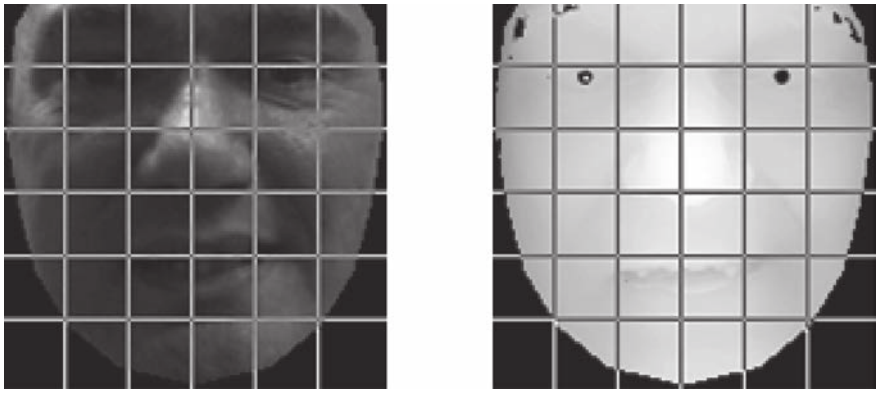

Fig. 2. Region division scheme.

3. Feature selection and decision level fusion

LCP histogram based features extracted from various sub-regions of facial texture and range images have different discriminative abilities to distinguish between genders and ethnicities, and thus present non-equal contributions to the final classification accuracy. Moreover, the manner for facial representation by using the division scheme tends to incur a very high dimensional feature space leading to expensive time and memory cost, and may even give rise to the problem of curse of dimensionality. To overcome these shortcomings, the step of feature selection is necessary. In this study, the Adaboost algorithm is exploited to select a compact subset of features from the whole feature set. The reason to use Adaboost for feature selection is that it is able to train a strong classifier while selecting features, and there is thus no need to re-train a classifier in the process of label prediction. For histogram based features, we can either treat each bin in the histograms or an individual histogram as a single feature. Considering that the differences between genders or races lie in discrimination of certain local circular patterns rather than all of them, therefore, we apply Adaboost to choose a set of discriminative bins as in [37,38] that also employ it to select LBP-like features.

Adaboost, originally proposed by Freund and Schapire [39], iteratively selects a small number of weak classifiers whose performances are just better than random guess, and boosts them into a strong classifier. Viola et al. [40] employed a variant of Adaboost to do face detection, and proposed the first real-time face detection algorithm. A distribution on the training samples is maintained, and in each iteration, weak classifiers are trained based on each feature according to the distribution. The classifier with the lowest weighted error is selected, so the corresponding feature is chosen in this iteration. We make use of Viola's variant of Adaboost to select a subset of histogram bins for gender and ethnicity classification. Details of the Adaboost can be found in [40], but in order to maintain consistence of this paper, the algorithm is posted in Table 1.

The weak classifier $h_j(x)$ is defined as

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) > p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where the parity p_j controls the direction of the inequality between feature f_j and the threshold θ_j. The threshold θ_j is calculated as the average of weighted centers of positive and negative samples' features.

Table 1
The Adaboost algorithm for LCP based feature selection (Ref [40]).

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where x_i stands for a sample, and $y_i = 0$, 1 for negative and positive examples respectively.
- Initialize weights $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0$, 1 respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$:
  1. Normalize the weights,
     $\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{j=1}^{n} \omega_{t,j}}$
     so that ω_t is a probability distribution.
  2. For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to ω_t, $\epsilon_j = \sum_i \omega_i \| h_j(x_i) - y_i \|$.
  3. Choose the classifier, h_t, with the lowest error ϵ_t.
  4. Update the weights:
     $\omega_{t+1,i} = \omega_{t,i} \beta_t^{1-e_i}$
     where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:
  $h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$
  where $\alpha_t = log\frac{1}{\beta_t}$.

Adaboost is performed for both texture and range image features, and results in two strong classifiers $h(x)$ and $h'(x)$, one for each modality. As shown in Table 1, these two classifiers are defined as below:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$h'(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha'_t h'_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha'_t \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

During testing, decision level fusion is performed. With the output of $h(x)$ and $h'(x)$, the final decision is made according to

$$H(x) = \begin{cases} 1 & \sum_{t=1}^{T} (\alpha_t h_t(x) + \alpha'_t h'_t(x)) \\ & \geq \frac{1}{2} \sum_{t=1}^{T} (\alpha_t + \alpha'_t) \\ 0 & \text{otherwise}. \end{cases} \tag{10}$$

Even though, Adaboost was proposed to solve the two-class problem as the one of gender classification, i.e. distinguishing male from female, it can also deal with the multi-class problem, e.g. ethnicity classification, by training a strong classifier for every two classes. In the test phase, the probe is decided by each of the strong classifiers learnt in the training phase, and is predicted with the label of class which has the proximal similarity measurement. Recently, some variants of Adaboost have been investigated to classify multiple classes [41,42], and they can be exploited as well.

4. Experimental results

In order to evaluate the effectiveness of the proposed LCP approach in the task of multi-modal facial gender and ethnicity classification, experiments are carried out on the FRGC v2.0 and BU-3DFE databases. We introduce the datasets and corresponding results subsequently.

4.1. Experiments on FRGC v2.0

FRGC v2.0 [43] is one of the most comprehensive datasets publicly available for 3D face analysis. It contains 4007 textured 3D face models of 466 subjects, and each face model is made up of a 3D point-cloud and its 2D texture counterpart. Among the subjects, 22% are Asian, 68% are white, and 10% are others. While 57% are male and 43% are female, with the age distribution: 65% 18–22 years old, 18% 23–27 and 17% 28 years or over. The database was collected during the 2003–2004 academic year, and contains time and illumination variations. Expressions such as “Neutral”, “Happiness”, “Surprise”, “Disgust”, “Sadness”, and “Other” are included in the database as well.

Gender classification is a typical binary classification problem, but the one for ethnicity classification is generally not the case. However, since the distribution of 3D face samples in current public databases (including FRGC v2.0) with ethnicity labels available is generally unbalanced, we also treat the task of ethnicity classification as a binary classification problem, in the same way as the previous studies do [19,17]. Asian and white subjects are thus chosen from the entire database for both gender and ethnicity classification, and in totality there are 3676 face samples belonging to 319 white and 99 Asian people. Even though all face samples in the FRGC v2.0 dataset are nearly frontal, Iteratively Closet Point (ICP) [44] is adopted to align the face model to the reference that is pre-defined, in order to control the error caused by slight pose changes. We then extract a pair of registered facial range and texture image from the aligned face model and all of the facial texture and range images are normalized so that the outer corners of two

eyes have a fixed distance of 100 pixels. An average mask is further used to eliminate non-face regions and segment face out, and finally images are normalized to the size of 140×140 pixels. Examples of normalized face images are shown in Fig. 2.

We design four experiments: the first is to test the performance of the Euclidean distance (L2) and the city block (L1) distance in clustering based quantization; the second is to analyze the robustness of the LCP descriptor to the LBP based one; the third is to evaluate the effectiveness of the proposed approach to gender and ethnicity classification in the modality of 2D, 3D and their combination; and the last one is to make the comparison with the state of the art techniques.

4.1.1. Performance of L1 and L2 distance in clustering based quantization

In order to evaluate the performance of L1 and L2 distance in clustering based quantization adopted in LCP, we randomly select 20 textured 3D face models from the FRGC database, and extract the features of local circular patterns $LCP_{2,8}$ from both the texture and shape modality as plotted in Fig. 3(a) and (d). The clustering results using L2 distance and L1 distance with the same number of clusters (59, identical to the number of bins in the LBP descriptor with 8 neighboring pixels) are shown in Fig. 3. In Fig. 3, the x-axis denotes the index of each element in an LCP code starting from the left-top position as shown in Fig. 1(a), and the y-axis displays the exact value of each element. In this case, the number of pixels around a central pixel is set at 8, thereby leading to 8 elements in an LCP code, and their values vary in the range of $[-255, 255]$.

We observe the difference between distributions of texture and shape data, and find out that the texture data are more concentrated than the shape data, therefore using the same quantization parameter like the traditional LBP to deal with these different distributions is obviously unsatisfied. In contrast, both the clustering results of L2 and L1 reflect the distribution difference much better. Meanwhile, the results obtained with L2 distance and L1 distance are different: the results of L2 distance are more evenly spaced than the ones of L1 distance, while L1 distance results focus on the concentrations of the data and pay less attention to the outlier of the distribution, likely leading to better performance in the step of gender and ethnicity classification (we show these accuracies subsequently).

Furthermore, we compare the computational cost for each distance metric, i.e. L1 and L2. Experiments are carried out on a PC with Intel Core i3 CPU using Matlab implementation. With 314,960 training local circular patterns, the computational time taken by K-means clustering with L2 distance repeated 10 times is 12,233 s, while that for L1 distance is only 600 s. L1 distance is thus computationally more efficient than L2 distance as well.

4.1.2. Analysis on robustness to noise of LCP

The performance under noise influence of clustering based quantization in LCP vs. binary quantization in LBP is also evaluated. Gaussian random noises with deviation of 2 are added on both facial texture and range images. Then local circular patterns $LCP_{2,8}$ are extracted before and after adding noise, and quantized using K-means clustering with L1 and L2 distances respectively. The number of clusters is set to 59 so as to achieve a fair comparison with the $LBP^{u2}_{2,8}$ operator of the same dimensionality. Histograms of quantization labels are constructed for the entire image. Fig. 4 shows an example, from which we can see that the influence of noise is more serious to the LBP based histograms than the LCP based one.

In order to quantitatively measure the difference between histograms, the distance $Diff(B, A)$ of histograms extracted before (B) and after (A) noise is added is calculated as below.

$$Diff(B,A) = \sum_i |B_i - A_i| \tag{11}$$

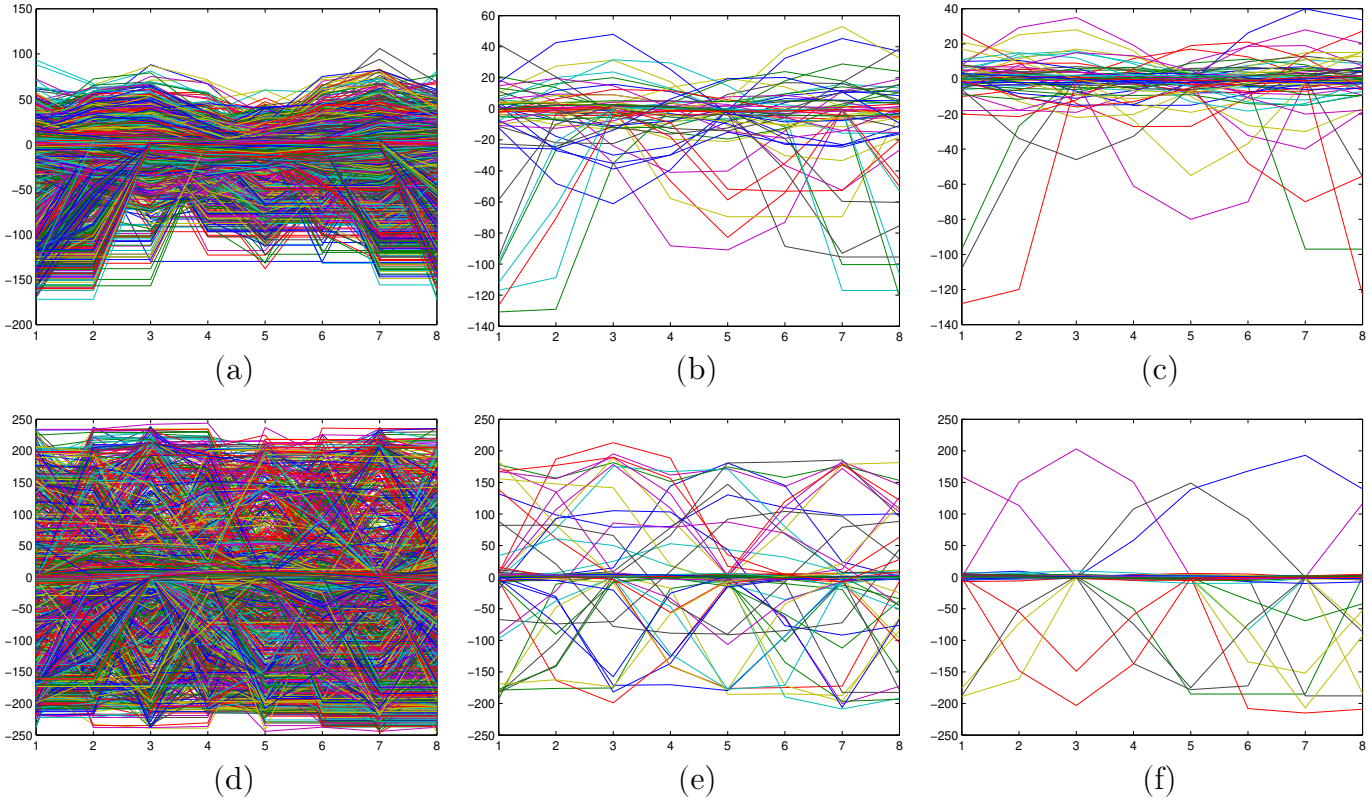

Fig. 3. The clustering results on the FRGC v2.0 database in the texture and shape modality respectively: (a) training texture data for clustering, (b) clustering result of texture data using L2 distance, (c) clustering result of texture data using L1 distance, (d) training shape data for clustering, (e) clustering result of shape data using L2 distance, and (f) clustering result of shape data using L1 distance.

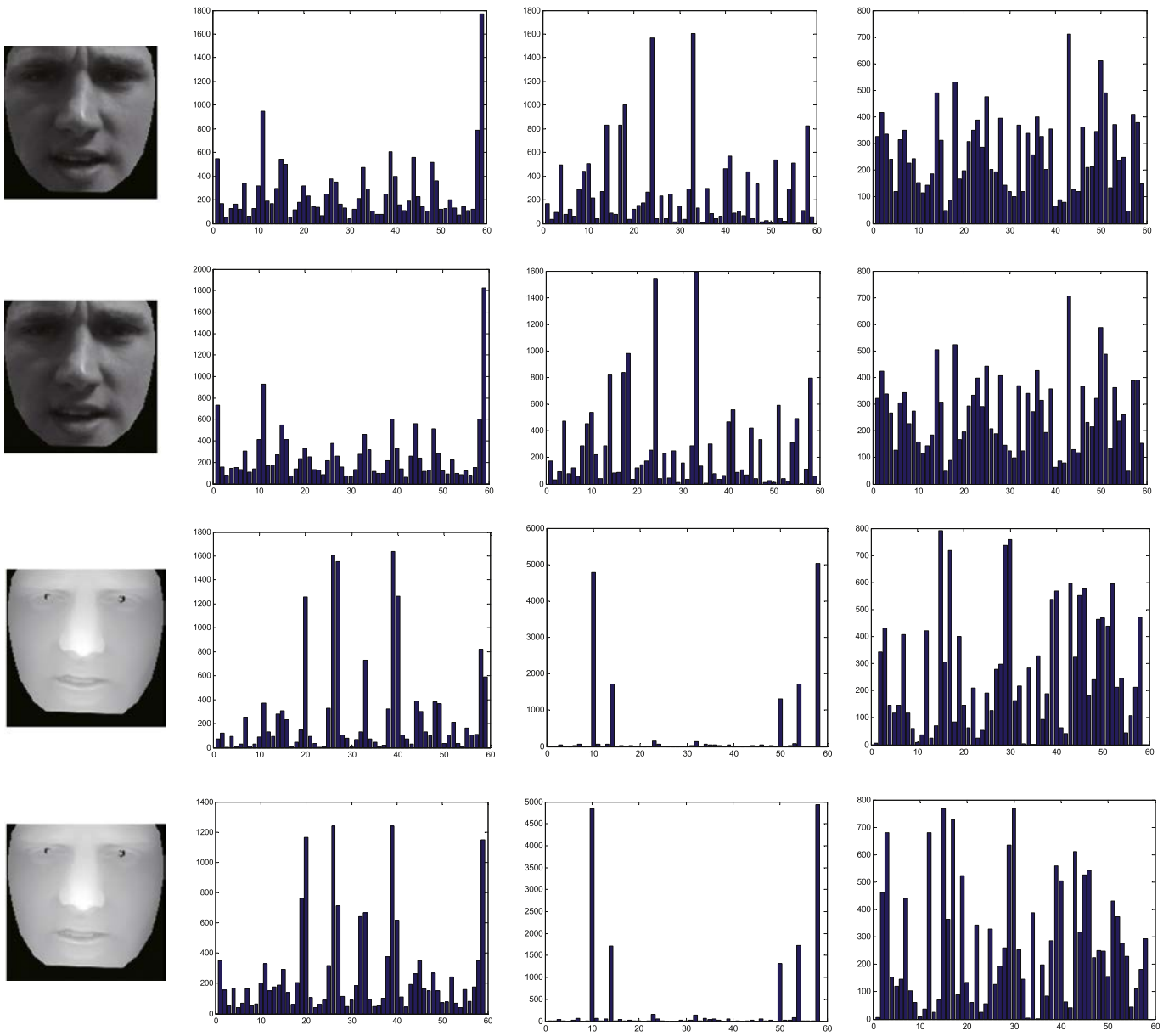

Fig. 4. Histograms extracted before (1st row and 3rd row) and after (2nd row and 4th row) noise adding using the samples from FRGC v2.0. The second column shows LBP histograms, in the third column are histograms extracted with L2 distance, histograms extracted with L1 distance are plotted in the fourth column.

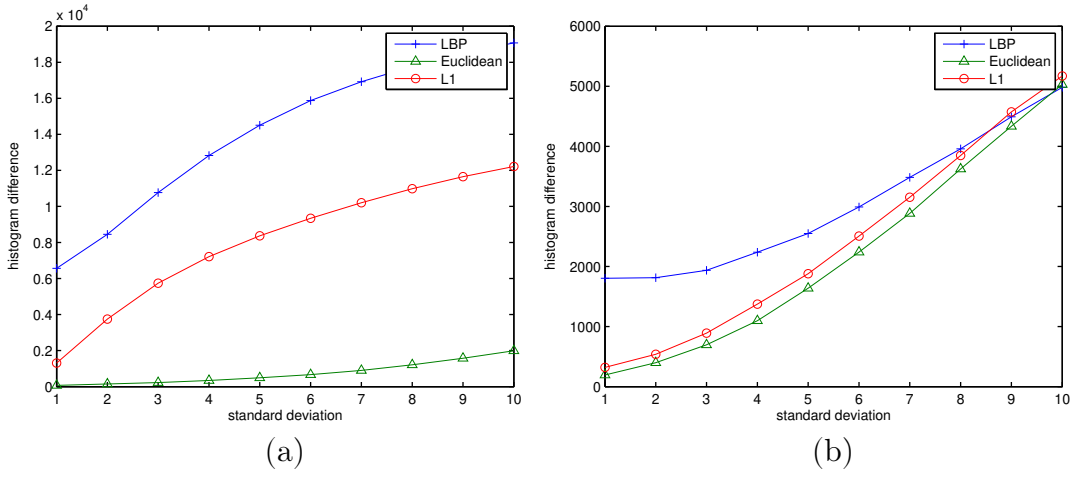

Fig. 5. The difference between histograms extracted before and after noise adding using the samples from FRGC v2.0: (a) for the shape modality and (b) for the texture modality.

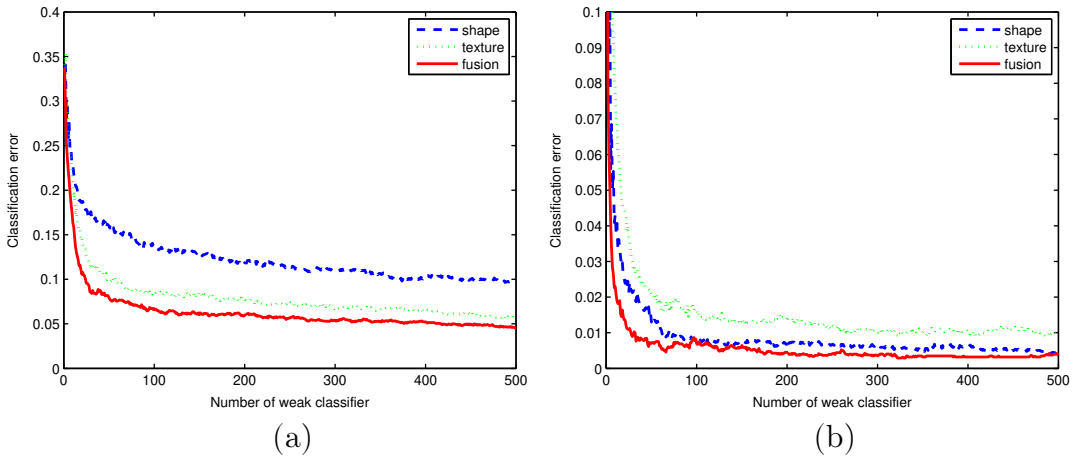

Fig. 6. Classification results of LCP achieved on the FRGC v2.0 database: (a) gender classification and (b) ethnicity classification.

where B and A are the two histograms to be compared, and B_i and A_i are the ith bin value of B and A. 20 samples are randomly selected from the FRGC v2.0 database as in the previous experiment, and Gaussian random noises with deviations vary from 1 to 10 and are added on range and texture images. Fig. 5 shows the average difference of histograms for LBP, LCP-L2, and LCP-L1. From the comparison, we can conclude that in both modalities, the clustering based quantization (LCP-L2, LCP-L1) outperforms binary quantization (LBP) under noise influence. Clustering with L2 distance performs better than L1 distance.

4.1.3. Gender and ethnicity classification with LCP

This experiment tests the performance of K-means clustering quantization based local circular patterns (LCP) in both tasks of gender and ethnicity classification. For multi-modal facial representation, $LCP_{8,1}$, $LCP_{8,2}$ and $LCP_{8,3}$ are used to extract features from the facial texture and range images, and quantized using K-means clustering. L1 distance is utilized in clustering due to its efficiency. All 2D texture and 3D range images are divided into 6×6 rectangular facial regions as shown in Fig. 2. As a result, for each modality of a face, three LCP histograms are extracted, and they are concatenated again to construct the final description of this modality. Through extracting LCP based features using multi-resolution filters and calculating histograms hierarchically, we can extensively find those distinctive features to represent gender and ethnicity related texture and shape variations.

As we defined in Section 2.2, N is the number of local circular patterns used to compute cluster centers. Actually, in our experiments, N corresponds to the number of 2D or 3D facial images randomly selected, since we have to ensure that these patterns are in an even distribution for face analysis. Therefore once a facial image is chosen, these local circular patterns of all pixels are used in the K-means clustering. We further vary the number of facial images for training and observe its impact on classification performance on the FRGC v2.0 dataset. In gender classification, when this number reaches about 50 and 60 for 2D and 3D modality respectively, their accuracies remain stable; while in ethnicity classification, this number is around 30 and 40. Meanwhile, we cannot set this number too large in order to avoid overfitting. As a result, in the following experiments, we set it at 60 so that it fits the two modalities in both tasks.

A 10-fold cross validation is adopted to evaluate the performance of the proposed approach, in which the database is randomly partitioned into 10 folds. Experiments are carried out 10 times, and each time 9 folds are exploited as the training set, and the remaining 1 fold as the testing set. Thus, each fold is tested once. We ensure that each subject is only assigned to one fold, so that the classification is person independent.

Fig. 6 shows the results of gender and ethnicity classification. As we can see, for gender classification texture features outperform shape features, while for ethnicity classification shape features perform better than texture features. In both tasks of gender and ethnicity classification, performance in either of the single modality is enhanced by combining shape and texture modalities. The classification errors achieved by fusion of these two modalities in the experiments of gender and ethnicity classification are 4.55% and 0.37% respectively. The confusion matrixes for gender and ethnicity classification are displayed in Tables 2 and 3.

We then analyze the results of L1 and L2 distance based K-means clustering quantization by comparing their accuracies in both classification tasks, i.e. gender and ethnicity. Fig. 7 shows the curves of classification errors vs. the number of weak classifier, and Table 4 shows the

Table 2
Confusion matrix of gender classification using LCP-L1 on the FRGC v2.0 dataset, and each item is depicted in the form of (average, standard deviation).

	Shape		Texture		Fusion	
	Male	Female	Male	Female	Male	Female
Male	(0.9097, 0.0478)	(0.0903, 0.0478)	(0.9476, 0.0345)	(0.0524, 0.0345)	(0.9596, 0.0324)	(0.0404, 0.0324)
Female	(0.1017, 0.0326)	(0.8983, 0.0326)	(0.0579, 0.0512)	(0.9421, 0.0512)	(0.0509, 0.0473)	(0.9491, 0.0473)

Table 3
Confusion matrix of ethnicity classification using LCP-L1 on the FRGC v2.0 dataset, and each item is depicted in the form of (average, standard deviation).

	Shape		Texture		Fusion	
	White	Asian	White	Asian	White	Asian
White	(0.9987, 0.0024)	(0.0013, 0.0024)	(0.9941, 0.0081)	(0.0059, 0.0081)	(0.9990, 0.0041)	(0.0010, 0.0041)
Asian	(0.0133, 0.0260)	(0.9867, 0.0260)	(0.0198, 0.0254)	(0.9802, 0.0254)	(0.0087, 0.0220)	(0.9913, 0.0220)

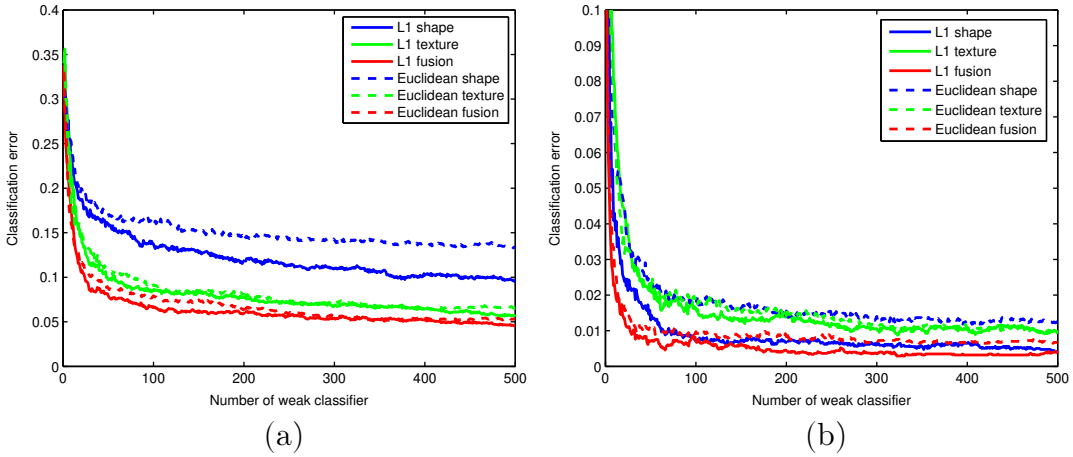

Fig. 7. Comparison of classification results between L2 and L1 distance of the LCP descriptor on the FRGC v2.0 database: (a) gender classification and (b) ethnicity classification.

comparison of classification errors between L1 and L2 distances. From these results we can see that L1 distance generally performs better than L2 distance, especially for the shape modality, while for texture modality the two distance metrics perform similarly.

4.1.4. Comparison with state of the art

In this experiment, we compare the performance of LCP with related local descriptors, i.e. LBP and one of its best variants, namely Complete LBP (CLBP) [28]. Using the same parameters in neighborhood setting, i.e. combining the histograms extracted in the neighborhood of (8,1), (8,2), and (8,3), the uniform LBP results in 59 different values while CLBP provides 200 different values for each histogram. Fig. 8 and Fig. 9 compare the classification error with respect to the number of weak classifier curves among these three methods, and Table 5 compares the classification error achieved by these methods. We can see from the results that the proposed method (LCP) consistently outperforms the LBP and Complete LBP methods, which demonstrate the superiority of clustering-based quantization of local circular patterns over binary quantization scheme used by LBP and CLBP.

Meanwhile, we compare the performance of the proposed approach with the ones of the state of the art techniques, which also concentrate on classifying gender and ethnicity using both the modality of 2D texture and 3D shape of human faces. Table 6 summarizes the comparison to highly related tasks. Although we carry out experiments with significantly more scans and more subjects, the accuracies of both gender and ethnicity classification are higher than the ones in [19,45]. The performance of gender classification is slightly lower than that of the work [46], while it should be noted that Huynh et al. [46] make use of uniform LBP features and Gradient-LBP features (a special case of CLBP) extracted from facial texture and range images respectively, and these features prove inferior to the proposed LCP features in our experiments. Furthermore, their result is based on 1149 pairs of facial range and gray images of 105 subjects, and the experiment is performed only once with half samples for training and half for testing. In our work, we carry out the experiment using 10-fold cross validation, where 3676 textured 3D face models of 418 subjects are involved.

Furthermore, we also list the work [17] that only makes use of the 3D modality in Table 6 since it exploits the same experimental protocol as we do. If regarding the shape information, LCP achieves comparable results as [17] does in ethnicity classification and it does not perform as good as [17] in gender classification, but when our system combines the clues of texture and shape, the accuracies in both the tasks are improved, which surpass the ones in [17]. Such a fact highlights the advantage of multimodal facial gender and ethnicity classification over the single modality based one. Additionally, [17] employs the 3D face recognition system (URxD) to measure the similarity of faces, holding a pipeline of deformed model based 3D surface registration and Haar wavelet decomposition as well as steerable pyramid transform based feature extraction, which is computationally expensive. In contrast, our system tends to be more efficient.

4.2. Experiments on BU-3DFE

BU-3DFE [47] is also one of the most popular databases in 3D face analysis, especially for 3D facial expression recognition. It contains 100 subjects among which 56 are female and 44 are male, ranging from 18 to 70 years old. All individuals are asked to perform six prototypic expressions. Each includes four levels of intensities, and there are hence 25 instant 3D expression models for each subject (plus one model with a neutral expression), leading to 2500 models in total. Median filter is utilized to remove spikes and cubic interpolation is adopted to fill holes. We employ ICP to align the face model to the pre-selected reference to correct possible pose variations. The registered facial range and texture image are then generated from each aligned face model and all of the facial texture and range images are normalized to the size of 140×140 pixels as in FRGC v2.0.

For gender classification, all 2500 3D face models belonging to these 100 persons are used in our experiments. While for ethnicity classification, due to the imbalance distribution of different races (51 Whites, 24 East-Asians, 9 Blacks, 8 Hispanic–Latinos, 6 Indians, and 2 Middle-East Asians), we exploit 1875 face models of Whites and East-Asians, as a binary classification problem. The settings of probe and gallery set in both tasks are the same as those of FRGC v2.0 stated in Section 4.1.3. For each

Table 4
Performance comparison between L1 distance and L2 distance of the LCP descriptor on the FRGC v2.0 database, and each item is depicted in the form of (average, standard deviation).

	Ethnicity			Gender		
	Shape	Texture	Fusion	Shape	Texture	Fusion
L1	(0.0042, 0.0058)	(0.0096, 0.0081)	(0.0037, 0.0051)	(0.0949, 0.0330)	(0.0557, 0.0262)	(0.0455, 0.0272)
L2	(0.0122, 0.0099)	(0.0094, 0.0092)	(0.0067, 0.0083)	(0.1344, 0.0380)	(0.0654, 0.0327)	(0.0536, 0.0279)

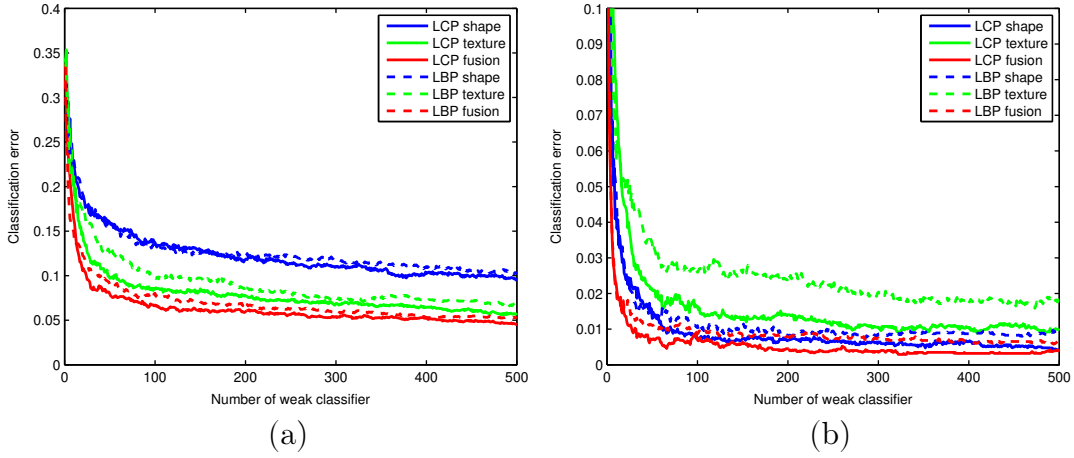

Fig. 8. Comparison of classification results between LCP and LBP on the FRGC v2.0 database: in (a) gender classification and (b) ethnicity classification.

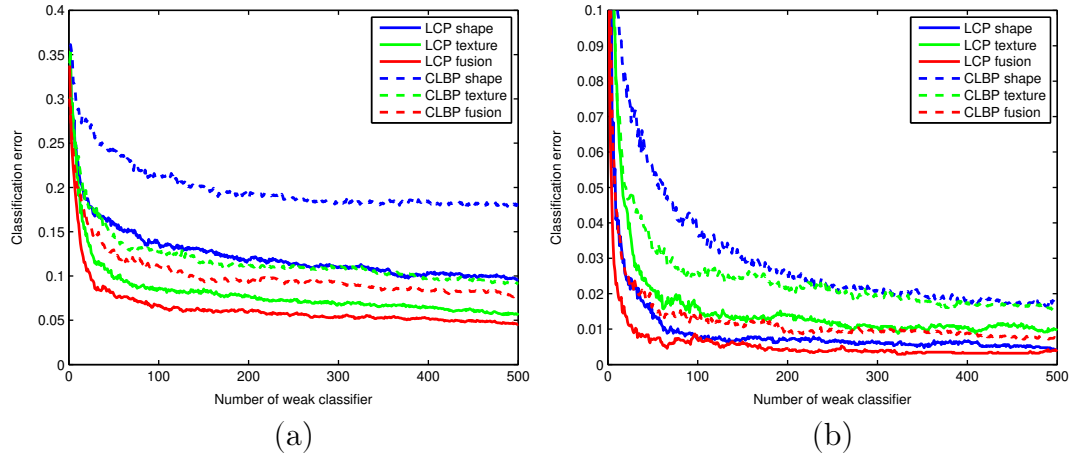

Fig. 9. Comparison of classification results between LCP and CLBP on the FRGC v2.0 database: in (a) gender classification and (b) ethnicity classification.

experiment, we make use of 10-fold cross validation and calculate the average performance.

4.2.1. Discussing the number of cluster centers in LCP

As we mentioned in Section 2.2, one of the key issues associated with K-means clustering is the number of clusters. This number controls the balance between descriptive power and sensitivity to noise. We experimentally evaluate this factor in the task of gender and ethnicity classification subsequently.

Taking the $LCP_{1,8}$ operator as an example, we increase the number of cluster centers from 20 to 100 at an interval of 5, and discover that the best performance of texture, shape or their combination is achieved in the range of [50, 70] for L1 and L2 distances in both tasks (as depicted in Fig. 10–Fig. 12), which coincidentally accords with the number (59) previously assigned for fair comparison with LBP.

4.2.2. Comparison with the approaches in the Literature

In this experiment, except LBP and its two variants, namely LTP [26] and CLBP [28], we also compare LCP with the features in [19,48,49]. [19] applies the holistic feature which is the raw pixels of a number of patch cropped from the facial image (denoted as "Grid" in Table 7). [48] and [49] both focus on texture classification, and we discuss them since they both make use of K-means clustering to learn the vocabulary of local pixel patterns. However, in LCP, we define the pattern as the gray value difference between the central pixel and its neighboring ones within the patch, rather than the original gray values [48] or their Random Projection (RP) [49] (denoted as "Pixel" and "RP" respectively in the following).

For LBP, LTP, LCP, and CLBP, as in FRGC v2.0, we combine the results of different neighborhood settings, i.e. (1, 8), (2, 8), and (3, 8), and divide the facial texture and range images into 6×6 regions. For "Grid", we directly inherit the parameters set in [19] that the number of patches

Table 5
Performance comparisons among LCP, LBP, and Complete LBP on the FRGC v2.0 dataset, and each item is depicted in the form of (average, standard deviation).

	Ethnicity			Gender		
	Shape	Texture	Fusion	Shape	Texture	Fusion
LCP	(0.0042, 0.0058)	(0.0096, 0.0081)	(0.0037, 0.0051)	(0.0949, 0.0330)	(0.0557, 0.0262)	(0.0455, 0.0272)
LBP	(0.0088, 0.0081)	(0.0182, 0.0105)	(0.0063, 0.0082)	(0.1034, 0.0346)	(0.0686, 0.0342)	(0.0524, 0.0308)
CLBP	(0.0177, 0.0114)	(0.0153, 0.0078)	(0.0074, 0.0086)	(0.1807, 0.0481)	(0.0909, 0.0367)	(0.0791, 0.0403)

Table 6
Performance comparison with those of the state of the art techniques of multi-modal facial gender and ethnicity classification on the FRGC v2.0 database (* indicates an exception that only makes use of the 3D modality, and the figures in bold are the best ones in individual tasks).

Approach	Sub. num.	Protocol	Gender	Ethnicity
Lu et al. [19]	376 Sub.& 1240 scans	10-fold C.-V.	91.00% ± 0.03	98.00% ± 0.16
Wu et al. [45]	260 (200 vs. 60) scans	6 Times	93.60% ± 0.04	–
Huynh et al. [46]	105 Sub.& 1149 scans	1 Time	96.70%	–
Toderici et al. [17]*	418 Sub.& ,676 scans	10-fold C.-V.	≈93.50%	≈99.50%
Our Method	418 Sub.& 3676 scans	10-fold C.-V.	**95.50% ± 0.03**	**99.60% ± 0.01**

is 80 (8×10) and the size of each patch is 8×8 pixels. For "Pixel" and "RP", to make a fair comparison with the LBP family, we set the patch size at 7×7, approximately equivalent to the combination of three neighborhood sizes in LBP, LTP, LCP, and CLBP, and employ their division scheme, i.e. 6×6 blocks for each face image. Additionally, in RP, we project the patch in an 8-dimensional PCA subspace for clustering.

From Table 7, we can see that:

- The performance of LCP-L1 and LCP-L2 in both gender and ethnicity classification is better than that of its counterparts in LBP family, i.e. uniform LBP, LTP, CLBP, on either of the single texture or shape modality as well as their combination, except the case in shape based ethnicity classification where the accuracy of LCP is only 0.29% below that of CLBP (still comparable). This fact clearly indicates that LCP is an effective improvement to the LBP methodology. Meanwhile, in LCP, LCP-L1 always performs LCP-L2, showing that the L1 distance is a better choice to learn LCP code than the L2 distance does.
- Regarding on Pixel-L1 and Pixel-L2 or RP-L1 and RP-L2 which apply the K-means clustering technique (using different distances) to learn local descriptors from original gray level values within a patch [48] or from their random projection [49], the LCP descriptor is competitive as well. Even though the results of LCP are slightly inferior to those of Pixel and RP (using L1 distance) based on texture clues in gender classification, the results of LCP in other tasks (including

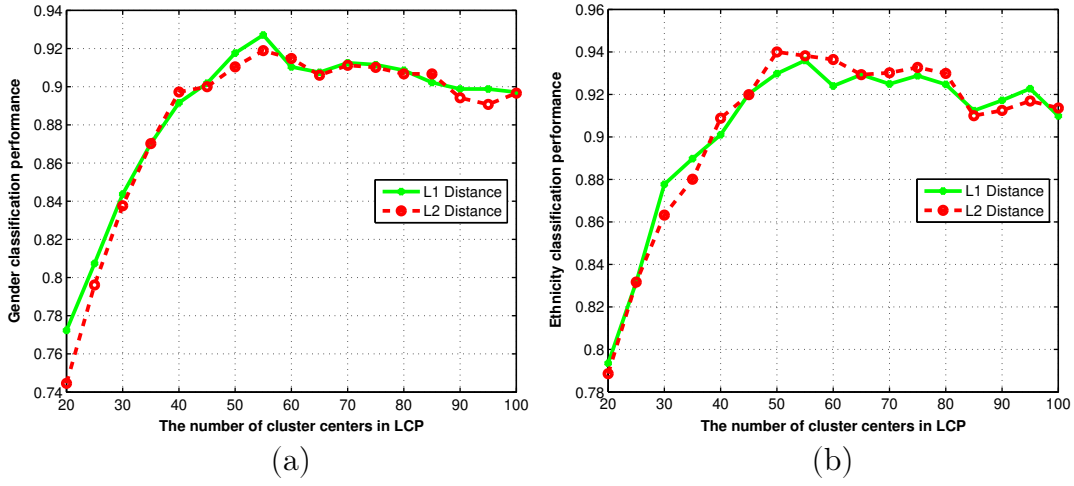

Fig. 10. Performance based on texture with regard to the number of cluster centers in $LCP_{1,8}$ in (a) gender classification and (b) ethnicity classification on the BU-3DFE dataset.

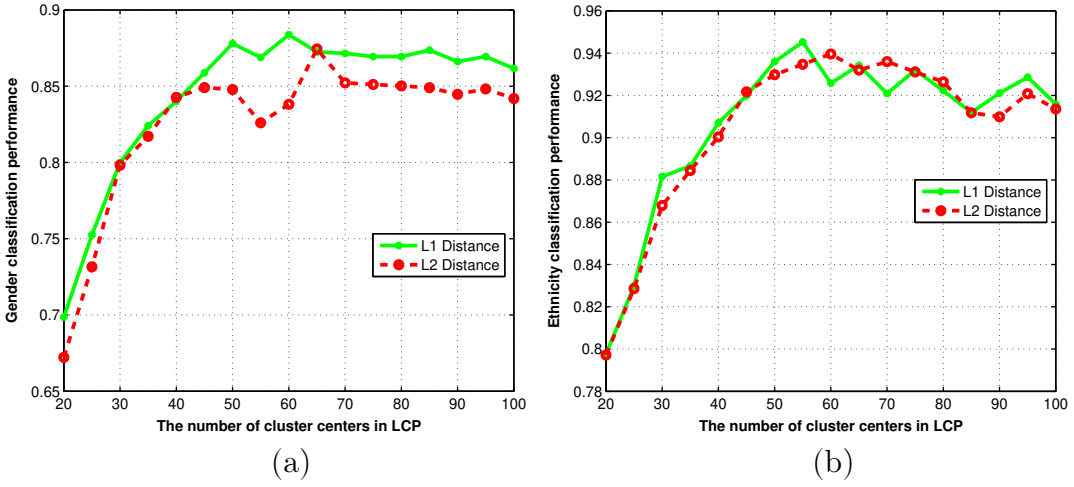

Fig. 11. Performance based on shape with regard to the number of cluster centers in $LCP_{1,8}$ in (a) gender classification and (b) ethnicity classification on the BU-3DFE dataset.

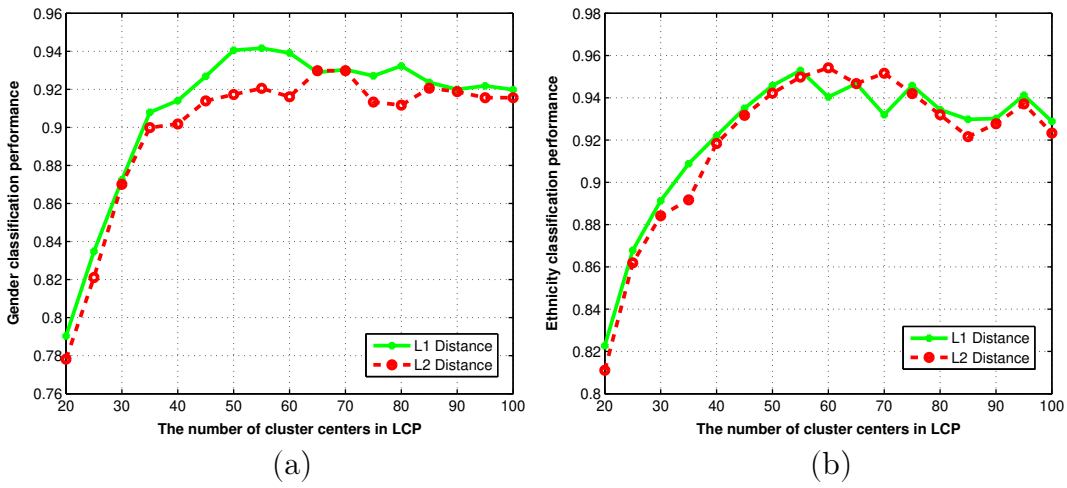

Fig. 12. Performance based on multi-modal combination with regard to the number of cluster centers in $LCP_{1,8}$ in (a) gender classification and (b) ethnicity classification on the BU-3DFE dataset.

shape-, fusion-based gender classification as well as texture-, shape-, and fusion-based ethnicity classification), are significantly superior to the ones of Pixel and RP, especially in the 3D modality. The reason mainly lies in that the facial range image is generally smooth and thus lacks discrimination, while the differences between neighboring pixels better highlight its details that are critical in classification than the original pixels [48] or their random projection [49]. It demonstrates the effectiveness of LCP for such issues.

- For all these methods discussed in the table, their classification accuracies based on the fusion of texture and shape cues are better than the corresponding ones using either of the single modality, illustrating that combining information conveyed in two modalities improves the performance in facial gender and ethnicity classification.

4.3. Time complexity evaluation

The K-means clustering based quantization in LCP is time consuming. However, it should be noted that this stage is carried out offline, and these cluster centers need to be generated only once during the training process. In online feature extraction, the only difference between LCP and LBP lies in that LCP calculates certain distance between a given local circular pattern and the pre-computed cluster centers and chooses the minimum one for quantization; while LBP makes use of binary quantization. The time cost additional to LBP is thus the distance calculation with all cluster centers. The more the cluster centers are, the higher the time cost is. In our experiments, there are 59 cluster centers, and based on C++ implementation, the average time cost of LBP, LCP-L1, and LCP-L2 is 3.33 ms, 9.76 ms, and 9.72 ms, respectively. Such computational complexity is generally under control in efficient face analysis applications.

5. Conclusion

In this paper, we present an effective and efficient approach on face based gender and ethnicity classification by combining both boosted local texture and shape features extracted from 3D face models. The proposed method is in contrast to the existing ones that only make use of either modality of 2D texture or 3D shape of faces. To comprehensively represent the difference between different genders or ethnicities, a novel local descriptor, namely local circular patterns (LCP) is introduced. LCP improves the widely investigated local binary patterns (LBP) as well as its variants by replacing the binary quantization with a clustering based one, thereby resulting in higher discriminative power and better robustness to noise. Moreover, the Adaboost based feature selection process finds the most discriminative gender- and race-related features and assigns them with different weights to highlight their importance in classification, which not only further raises the performance but reduces the time and memory cost as well. The experimental results of gender and ethnicity classification achieved are up to 95.50% and 99.60% respectively on the FRGC v2.0 dataset, and 95.60% and 97.42% respectively on the BU-3DFE dataset, which clearly demonstrate the advantages of the proposed method.

In future work, we will investigate possible solutions that are more effective to multi-modal facial gender and ethnicity classification, e.g.

Table 7
Comparison of approaches of the texture and shape modality as well as their combination using Adaboost in the task of gender and ethnicity classification on the BU-3DFE dataset. (The figures in bold are the best ones in individual tasks)

	Gender classification			Ethnicity classification		
	Texture	Shape	Fusion	Texture	Shape	Fusion
LBP	93.53% ± 0.03	87.35% ± 0.06	94.58% ± 0.03	95.07% ± 0.04	95.56% ± 0.04	96.89% ± 0.03
LTP	93.20% ± 0.01	87.44% ± 0.05	94.18% ± 0.03	94.91% ± 0.03	95.05% ± 0.04	96.62% ± 0.04
LCP-L1	94.18% ± 0.03	**90.76% ± 0.04**	**95.60% ± 0.03**	**97.31% ± 0.04**	96.33% ± 0.03	**97.42% ± 0.04**
LCP-L2	94.00% ± 0.04	89.09% ± 0.04	95.56% ± 0.03	97.13% ± 0.04	95.98% ± 0.03	97.22% ± 0.03
CLBP	93.82% ± 0.02	88.12% ± 0.04	94.91% ± 0.03	95.86% ± 0.04	**96.62% ± 0.02**	97.13% ± 0.04
Grid	81.27% ± 0.07	78.70% ± 0.07	85.60% ± 0.08	93.39% ± 0.10	87.87% ± 0.10	94.07% ± 0.05
Pixel-L1	94.36% ± 0.05	81.64% ± 0.04	95.45% ± 0.02	91.26% ± 0.05	91.70% ± 0.05	92.89% ± 0.01
Pixel-L2	91.64% ± 0.04	80.97% ± 0.05	93.82% ± 0.05	90.52% ± 0.06	91.26% ± 0.04	92.15% ± 0.01
RP-L1	**94.91% ± 0.04**	81.64% ± 0.03	95.27% ± 0.01	90.07% ± 0.06	92.30% ± 0.06	93.19% ± 0.02
RP-L2	94.18% ± 0.06	81.94% ± 0.01	94.42% ± 0.02	91.56% ± 0.06	91.41% ± 0.04	94.37% ± 0.02

based on the findings in [20], where general discriminant local face descriptors are learned.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61202237; the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20121102120016); the research program of State Key Laboratory of Software Development Environment (SKLSDE-2013ZX-31); the joint project by the LIA 2MCSI lab between the group of Ecoles Centrales and Beihang University; and the Fundamental Research Funds for the Central Universities.

References

[1] B.A. Golomb, D.T. Lawrence, T.J. Sejnowski, Sexnet: a neural network identifies sex from human faces, Conference on Advances in Neural Information Processing Systems, vol. 3, 1991, pp. 572–577.
[2] B. Moghaddam, M.-H. Yang, Gender classification with support vector machines, IEEE International Conference on Automatic Face and Gesture Recognition, 2000.
[3] X. Lu, A.K. Jain, Ethnicity identification from face images, SPIE International Symposium on Defense and Security: Biometric Technology for Human Identification, 2004, pp. 114–123.
[4] G. Shakhnarovich, P.A. Viola, B. Moghaddam, A unified learning framework for real time face detection and classification, IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
[5] S. Hosoi, E. Takikawa, M. Kawade, Ethnicity estimation with facial images, IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 195–200.
[6] S. Lao, M. Kawade, Vision-based face understanding technologies and their applications, Chinese Conference on Biometric Recognition, 2004, pp. 339–348.
[7] H.C. Lian, B.L. Lu, E. Talikawa, S. Hosoi, Gender recognition using a min–max modular support vector machine, International Conference on Natural Computation, 2005.
[8] H. Lu, H. Lin, Gender recognition using adaboosted feature, IEEE International Conference on Natural Computation, 2007.
[9] Z. Yang, H. Ai, Demographic classification with local binary patterns, International Conference on Biometrics, 2007, pp. 464–473.
[10] G. Guo, G. Mu, A study of large-scale ethnicity estimation with gender and age variations, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2010, pp. 79–86.
[11] R. Brunelli, T. Poggio, Hyberbf networks for gender classification, DARPA Image Understanding Workshop, 1992, pp. 311–314.
[12] A. Samal, V. Subramani, D. Marx, Analysis of sexual dimorphism in human face, J. Vis. Commun. Image Represent. 18 (6) (2007) 453–463.
[13] A.J. O'Toole, T. Vetter, N.F. Troje, H.H. Bulthoff, Sex classification is better with three-dimensional head structure than with image intensity information, Perception 26 (1) (1997) 75–84.
[14] X. Han, H. Ugail, I. Palmer, Gender classification based on 3D face geometry features using svm, International Conference on Cyberworlds, 2009, pp. 114–118.
[15] J. Wu, W. Smith, E. Hancock, M. Kawulok, Extracting gender discriminating features from facial needle-maps, International Conference on Image Processing, 2009, pp. 2449–2452.
[16] Y. Hu, J. Yan, P. Shi, A fusion-based method for 3D facial gender classification, International Conference on Computer and Automation, Engineering, 2010, pp. 369–372.
[17] G. Toderici, S. OMalley, G. Passalis, T. Theoharis, I. Kakadiaris, Ethnicity- and gender-based subject retrieval using 3-d face-recognition techniques, Int. J. Comput. Vis. 89 (2) (2010) 382–391.
[18] D. Huang, W. Ben Soltana, M. Ardabilian, Y. Wang, L. Chen, Textured 3d face recognition using biological vision-based facial representation and optimized weighted sum fusion, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2011, pp. 1–8.
[19] X. Lu, H. Chen, A. Jain, Multimodal facial gender and ethnicity identification, IEEE International Conference on, Biometrics, 2006, pp. 554–561.
[20] Z. Lei, M. Pietikainen, S.Z. Li, Learning discriminant face descriptor, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 289–302.
[21] N.-S. Vu, A. Caplier, Enhanced patterns of oriented edge magnitudes for face recognition and image matching, IEEE Trans. Image Process. 21 (3) (2012) 1352–1365.
[22] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 41 (6) (2011) 765–781.
[23] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[24] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.
[25] Y. Huang, Y. Wang, T. Tan, Combining statistics of geometrical and correlative features for 3d face recognition, British Machine Vision Conference, 2006, pp. 879–888.
[26] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2007, pp. 168–182.
[27] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using lbp variance (LBPV) with global matching, Pattern Recogn. 43 (3) (2010) 706–719.
[28] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, IEEE Trans. Image Process. 19 (6) (2010) 1657–1663.
[29] H. Yang, Y. Wang, A lbp-based face recognition method with hamming distance constraint, International Conference on Image and Graphics, 2007, pp. 645–649.
[30] S. Liao, M.W.K. Law, A.C.S. Chung, Dominant local binary patterns for texture classification, IEEE Trans. Image Process. 18 (5) (2009) 1107–1118.
[31] M. Meilă, The uniqueness of a good optimum for k-means, International Conference on Machine Learning, 2006, pp. 625–632.
[32] H. Kashima, J. Hu, B.K. Ray, M. Singh, K-means clustering of proportional data using L1 distance, IEEE International Conference on Pattern Recognition, 2008, pp. 1–4.
[33] D. Huang, M. Ardabilian, Y. Wang, L. Chen, 3-d face recognition using elbp-based facial description and local feature hybrid matching, IEEE Trans. Inf. Forensics Secur. 7 (5) (2012) 1551–1565.
[34] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, European Conference on Computer Vision, 2004, pp. 469–481.
[35] X.T.S. Yan, H. Wang, T.S. Huang, Exploring feature descriptors for face recognition, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007.
[36] J.K.C. Chan, K. Messer, Multi-scale local binary pattern histograms for face recognition, International Conference on Biometrics, 2007.
[37] C. Shan, T. Gritti, Learning discriminative lbp-histogram bins for facial expression recognition, British Machine Vision Conference, 2008.
[38] G. Zhao, M. Pietikainen, Principal appearance and motion from boosted spatiotemporal descriptors, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–8.
[39] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, European Conference on Computational Learning Theory, 1995, pp. 23–37.
[40] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 511–518.
[41] J. Zhu, H. Zou, S. Rosset, T. Hastie, Multi-class adaboost, Stat. Interface 2 (2009) 349–360.
[42] T.-H. Kim, D.-C. Park, D.-M. Woo, T. Jeong, S.-Y. Min, Multi-class classifier-based Adaboost algorithm, Intelligent Science and Intelligent Data Engineering, 2012, pp. 122–127.
[43] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 947–954.
[44] Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, Int. J. Comput. Vision 13 (2) (1994) 119–152.
[45] J. Wu, W. Smith, E. Hancock, Gender classification using shape from shading, International Conference on Image Analysis and Recognition, 2007, p. 499508.
[46] T. Huynh, R. Min, J.L. Dugelay, An efficient lbpbased descriptor for facial depth images applied to gender recognition using rgb-d face data, ACCV Workshop on Computer Vision with Local Binary Pattern Variants, 2012.
[47] J. Wang, L. Yin, X. Wei, Y. Sun, 3d facial expression recognition based on primitive surface feature distribution, vol. 2 (2006) 1399–1406.
[48] M. Varma, A. Zisserman, A statistical approach to material classification using image patch exemplars, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2009) 2032–2047.
[49] L. Liu, P. Fieguth, Texture classification from random features, IEEE Trans. Pattern Anal. Mach. Intell. 34 (3) (2012) 574–586.