International Journal on Artificial Intelligence Tools
Vol. 21, No. 5 (2012) 1250019 (24 pages)

DOI: 10.1142/S0218213012500194

RACE CLASSIFICATION FROM FACE IMAGES USING LOCAL DESCRIPTORS

GHULAM MUHAMMAD

Computer Engineering Department, College of Computer and Information Sciences
King Saud University, P. O. Box 51178, Riyadh 11543, Saudi Arabia
ghulam@ksu.edu.sa

MUHAMMAD HUSSAIN* and FATMAH ALENEZY†

Computer Science Department, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
**mhussain@ksu.edu.sa*
†fatmah.alenezy@gmail.com

GEORGE BEBIS

Department of Computer Science and Engineering
University of Nevada, Reno, USA
bebis@cse.unr.edu

ANWAR M. MIRZA

Computer Engineering Department, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
anwar.m.mirza@gmail.com

HATIM ABOALSAMH

Computer Science Department, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
hatim@ksu.edu.sa

Received 14 November 2011
Accepted 10 January 2012

This paper investigates and compares the performance of local descriptors for race classification from face images. Two powerful types of local descriptors have been considered in this study: Local Binary Patterns (LBP) and Weber Local Descriptors (WLD). First, we investigate the performance of LBP and WLD separately and experiment with different parameter values to optimize race classification. Second, we apply the Kruskal-Wallis feature selection algorithm to select a subset of more "discriminative" bins from the LBP and WLD histograms. Finally, we fuse LBP and WLD, both at the feature and score levels, to further improve race classification accuracy. For classification, we have considered the minimum distance classifier and experimented with three distance measures: City-block, Euclidean, and Chi-square. We have performed extensive experiments and comparisons using five race groups from the FERET database. Our experimental results indicate that (i) using the Kruskal-Wallis feature selection, (ii) fusing LBP with WLD at the feature level, and (iii) using the City-block distance for classification, outperforms LBP and WLD alone as well as methods based on holistic features such as Principal Component Analysis (PCA) and LBP or WLD (i.e., applied globally).

Keywords: Race recognition; face recognition; local binary pattern; Weber local descriptors.

1. Introduction

Face processing and recognition is a key biometric technology with a wide range of potential applications related to security and safety. Current research efforts in the field involve developing more accurate and robust algorithms for face detection and recognition as well as algorithms to classify faces with respect to gender, race, and age.[1,2] This information could be used to collect useful demographics but also enhance face recognition performance. There exists significant cognitive evidence supporting that humans utilize information from various visual cues for face recognition. It is well known, for example, that people are more accurate at recognizing faces of their own race than faces of other races (i.e., "other race effect").[3,4] Other studies have shown that humans judge the gender of adults and children using feature sets derived from the appropriate face age category, rather than applying features derived from another age category or from a combination of age categories.[5] Obviously, gender, race, and age information could be used to reduce the search space when matching unknown faces to a set of known faces but also to optimize face recognition algorithms using face categories.

In this paper, our interest is in improving race classification. Figure 1 illustrates the main steps involved in race classification. We assume that the face has already been detected by some prior step. First, pre-processing is applied to normalize the input face (e.g., with respect to size, location, orientation, and lighting). Then, feature extraction is performed to represent the face by a compact, more discriminative set of features. Finally, a decision mechanism is employed (e.g., classifier) to identify the race of the input face. The focus of this work is on investigating and comparing the performance of local features for race classification.

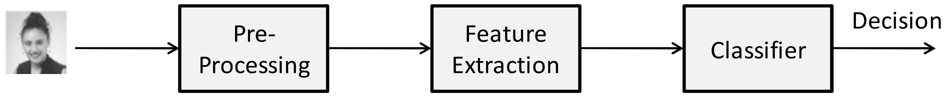


Fig. 1. Main steps involved in race classification.

Local feature-based methods have gained much attention over holistic based methods in the face recognition community for their robustness to illumination and pose variations.[15] Local feature-based methods employ features extracted from local regions of the face. Methods employing geometrical features such as distances and areas between facial features (e.g., eyes, nose, and mouth) fall into this category.[13] These methods, however, require accurate localization of various facial features which might not be easy in practice. As a result, methods relying on local features that do not rely on detecting facial features have gained a lot of popularity. Examples include LBP features, Histograms of Orientated Gradient (HoG) features, Scale Invariant Feature Transform (SIFT) features, WLD features, and Gabor wavelets.

LBP features were introduced by Ojala *et al.*[9] and used by Ahonen *et al.*[8] for face recognition. LBP features are computed by labeling the pixels of an image by thresholding the 3×3-neighbourhood of each pixel with the center value and considering the result as a binary number. The histogram of the labels is then used as a texture descriptor. Dalal and Triggs[14] use HOG features for human detection. They extract HOG features by counting occurrences of edge orientations in localized portions of an image. First, the image is divided into small cells and a histogram of edge orientations is computed for each cell. Then, the HOG descriptor is formed by combining the normalized histograms. SIFT features represent a class of highly distinctive local features proposed by Lowe.[11] They are invariant to image scale, rotation and illumination. Moreover, they are robust to occlusion, clutter or noise. SIFT features are extracted in four stages. First stage detects all interest keypoints that are invariant to scale and orientation in an image by searching over all scales and image locations. Second stage removes the unreliable keypoints which have low contrast (so it is sensitive to noise) or are poorly localized along an edge. The orientation is assigned to each keypoint based on image Gradient direction in the third stage. In the final stage, a local descriptor is found that is invariant to illumination and 3D view point. This descriptor is computed as a set of orientation histograms on 16×16 pixel neighborhoods. SIFT features have been used in many applications including face recognition.[12] Liu and Wechsler[10] review the basics of Gabor wavelets and describe the Gabor feature representation of an image. To derive a low-dimensional feature representation with enhanced discrimination power, they use eigenvalue selectivity constraints of the Enhance Fisher Model (EFM). WLD features[27] were inspired by Weber's law which has its roots to human perception. WLD features correspond to histograms of both differential excitations and orientations. Their computation is faster than that of SIFT features and slightly slower to that of LBP features. WLD features have been shown to outperform SIFT, LBP, and Gabor wavelet features on the task of texture classification.[27]

In this paper, we have experimented with LBP and WLD features as they are both "dense" local descriptors which can also be computed much faster compared to SIFT and Gabor wavelet features. Moreover, LBP and LWD features encode information differently. LPB represents an input image by building statistics on the local micro-pattern variation (e.g., bright/dark spots, edges and flat areas etc.). In contrast, WLD first computes the salient micro-patterns (i.e., differential excitation), and then builds statistics on these salient patterns along with the gradient orientation of the current point. First, we have performed experiments using LBP and WLD features separately to assess their performance in the context of race classification. To improve accuracy, we have applied feature selection using the Kruskal-Wallis algorithm and fusion, both at the feature and score levels. For classification, we have considered the minimum distance classifier and experimented with three different distance measures: City-block, Euclidean, and Chi-square. Extensive experiments using five race groups from the FERET database show that fusing LBP with WLD works better than using them independently and outperforms conventional holistic methods such as PCA.

2. Literature Review of Race Classification

Race classification using face images is relatively a new topic in computer vision. The "other race effect" has a strong influence on human perception. MacLin and Malpass[20] subjectively found that other race faces are encoded categorically and this categorization contributes to human perception. Phillips *et al.*[21] analyzed other-race effect on face recognition algorithms using the results of the 2006 Face Recognition Vendor Test (FRVT). They found that Western algorithms (i.e., developed by France, Germany and the United States research groups) recognized Caucasian faces more accurately than East Asian faces and East Asian algorithms (i.e., developed by China, Japan, and Korea research groups) recognized East Asian faces more accurately than Caucasian faces.

In Ref. 22, a two-class race classification problem was considered (i.e., Asian and non-Asian) using multiscale LDA. Modeling each face with a single Gaussian, they constructed an ensemble by integrating the classification results using dot product. An overall accuracy of 96.3% was reported in classifying 132 Asian faces and 131 non-Asian faces. Hosoi *et al.*[23] designed an ethnicity estimation method using featured based on Gabor wavelets and retinal sampling; an SVM was used for classification. Three types of ethnic groups were considered: African, Asian, and European. They reported an overall accuracy of approximately 94%. Lu *et al.*[24] used 3D models of faces to infer ethnicity between two groups (Asian and non-Asian). Using normalized range an intensity images, they employed two SVMs, one for each modality, to infer ethnicity. The final decision was made by integrating the two SVM results. Using 918 face images from 282 non-Asian subjects and 322 face images from 94 Asian subjects, they reported accuracy close to 98%. Zhiguang and Haizhou[25] used LBP features for demographic classification, which includes ethnicity and/or race. The AdaBoost algorithm was used along with Chi-square distance metric to form a strong classifier. Experimental results on a two-class race classification problem (i.e., Asian vs non-Asian) confirm that LBP features are comparable to Haar (wavelets) like features. The error rate using Haar features and LBP features was 2.98% and 3.01%, respectively. However, the effect of different LBP parameters was not investigated in their study.

Gutta *et al.*[31] fed gray scale pixel values to an ensemble radial basis functions with inductive decision trees for ethnicity classification. They achieved 94% accuracy in a four class problem (i.e., Caucasian, Oriental, African, and Asian) using the FERET database. Guo and Mu[7] used a large database (i.e., MORPH-II) to investigate the effects of gender and age on ethnicity classification using biologically inspired features based on Gabor filters and SVM for classification. Zhang and Wang[33] used multi-scale, multi-ratio LBP features extracted from 2D texture and 3D range face images for race classification. Using two major race groups (i.e., White and Asia) from the FRGC v.2 database, they reported a 0.42% error rate. Duan *et al.*[34] used LDA based algebraic features and an elastic model based on geometric features to classify three minor Chinese ethnic groups: Tibetan, Uighur, and Zhuang. They reported a 79% accuracy using algebraic features and 90.95% accuracy using geometric features with kNN and C5.0 classifiers.

Manesh *et al.*[26] considered a two class ethnicity classification problem (i.e., Asian and non Asian) using an appearance-based method to determine the confidence of different facial regions using Support Vector Machines (SVM).[26] They reported a 0.0261% error rate with faces normalized using eye and mouth positions. The used face images collected from the FERET and CAS-PEAL databases. A real time face detection and classification (i.e., gender and ethnicity) framework was proposed in Ref. 35. Three different types of rectangular filters were used to extract features, and SVM and boosted classifiers were used to classify face images in two-classes (i.e., Asian and non Asian). The method achieved a 22.6% error rate. Table 1 summarizes previous works on race recognition in terms of database used, race groups, feature extraction method, and classifier.

Table 1. Summary of previous works on race classification from face images.

Ref.	Database	Race Groups	Features	Classifier
22	AsianPF01, NLPR, Yale, AR	Asian, Non-Asian	LDA	Cosine distance
23	HOIP Other	Asian, European, African	Gabor wavelet transform and retina sampling	SVM
25	FERET snapshot PIE	Asian, Non-Asian	LBP	AdaBoost and Chi square distance
31	FERET	Caucasian, Asian, Oriental, African	Raw gray scale values	Ensembles of radial basis function and SVM
7	MORPH-II	Selected: Black and White	Biologically inspired features	SVM
24	3D face databases: University of Notre Dame (UND), Michigan State University (MSU)	Asian, Non-Asian	Interpolated features	SVM
33	FRGC v2.0	White, Asian, Other	Multi-scale multi-ratio LBP	AdaBoost and Chi-square minimum distance
34	Database collected by authors	Three minority ethnic groups in the Chinese race: *Tibetan, Uighur, Zhuang*.	LDA-PCA, Geometric features	kNN, C5.0
26	CAS-PEAL FERET	Asian, Non-Asian	Gabor	SVM
35	Faces from WWW	Asian, Non-Asian	Rectangular features	Boosted classifier and SVM

From the above discussion, it can be noted that race recognition using face images is not fully discovered with the state of the art features, though a significant progress has been made in face recognition. Some attempts are made for only two or three class problems, which is relatively easier than many class problems. Also no feature selection methods are applied to race recognition problems. Therefore, in this paper, we introduce a new race recognition method that (a) involves state of the art local features, (b) utilizes feature selection technique, and (c) works on five race groups (classes).

3. Race Classification Using Local Features

An overview of the proposed race classification system is shown in Figure 2. Two types of features are extracted from normalized input face images: LPB and WLD. The Kruskal-Wallis feature selection method is applied on each set of features to select a subset of discriminative features. The resulted LBP and WLD histograms are fused together through concatenation to produce a more powerful set of features. The minimum distance classifier is employed for identifying the race of the input faces. We have experimented with various distance measures; our best results were obtained using the City-block distance.

Next, we present a brief review of the LBP and WLD features, the Kruskal-Wallis feature selection algorithm, the minimum distance classifier and the distance measures tested, and the fusion of LBP and WLD features.

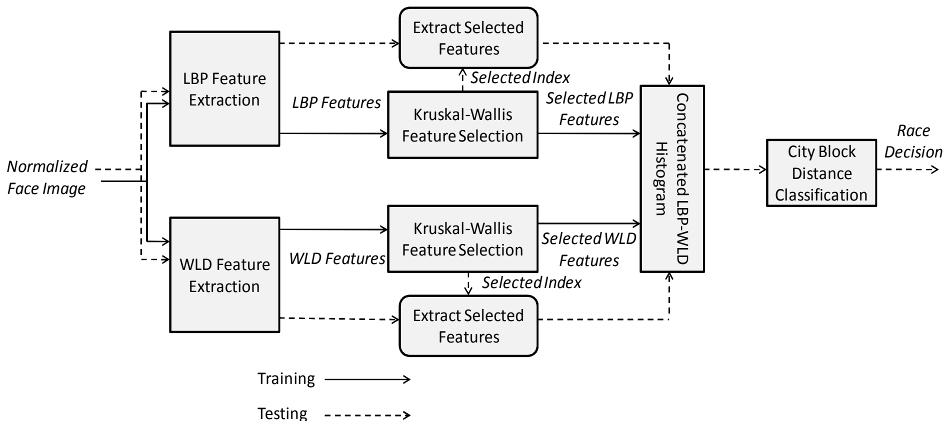


Fig. 2. Block diagram of the proposed race classification system.

3.1. *LBP*

Ahonen *et al.* introduced the method of LBP for face recognition[8] by using the original LBP operator that was proposed by Ojala *et al.*[9] LBP has been one of the most widely used and best performing texture descriptors in recent years. This operator labels the pixels of an image by thresholding the 3×3-neighbourhood of each pixel with the center

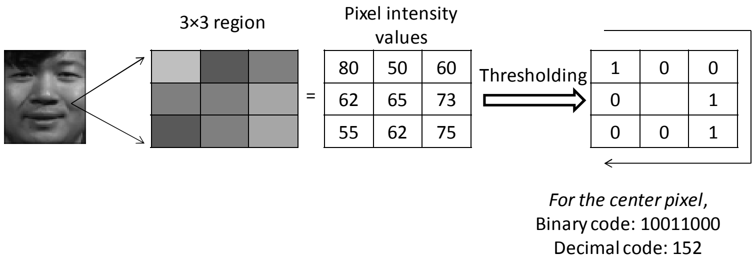


Fig. 3. Basic LBP operator.

value and considering the result as a binary number (see Fig. 3). Then the histogram of the labels can be used as a texture descriptor. We refer to this method as “basic” LBP. Later, the operator was extended to use a neighborhood of different size. Ahonen *et al.*[8] used the notation $\mathrm{LBP}^{\mathrm{u2}}_{\mathrm{P,R}}$ for a circular neighborhood where P denotes the number of sample points on a circle of radius R (see Fig. 4). The superscript ‘u2’ stands for using only uniform patterns and labeling all remaining patterns with a single label. Specifically, LBP is computed using the following equation:

$$LBP_{P,R} = \sum_{i=1}^{P-1} 2^i \times S(p_i - p_c) \tag{1}$$

where p_c is the center pixel and the thresholding operation is $S(p_i - p_c) = \begin{cases} 1, & p_i - p_c \geq 0 \\ 0, & p_i - p_c < 0 \end{cases}$.

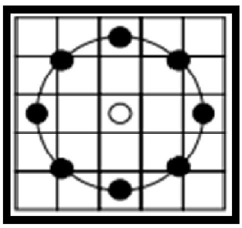


Fig. 4. Circular (P, R) = (8, 2) neigbourhood of LBP operator; here, P = 8 represent the number of samples on a circle of radius 2.

The histogram of the labeled image $f_l(x, y)$ is defined as:

$$H_i = \sum_{x,y} I\{f_l(x,y) = i\}, \quad i = 0,1,...,n-1 \tag{2}$$

where n is the number of different labels produced by the LBP operator and

$$I\{A\} = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false} \end{cases}$$

Each histogram contains information about facial micro-patterns like the distribution of edges, spots and flat areas over the whole image. However, one should keep local spatial information for efficient representation of face. So the image is divided into regions (i.e., blocks) $R_0, R_1, \ldots, R_{m-1}$ from which local binary patterns are extracted and a global feature histogram is constructed. The histogram represents the facial micro-patterns and their spatial location as follows:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\}, \quad i = 0,...,n-1 \text{ and } j = 0,...,m-1 \tag{3}$$

In using LBP for race classification, the following parameters were varied to optimize their performance: type of LBP operator and block size. The first parameter is $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{mapping}}$, where mapping corresponds to one of the following three mapping options: 'u2' for uniform LBP, 'ri' for rotation-invariant LBP, and 'riu2' for uniform rotation-invariant LBP; 0 is used for no mapping. LBP is called uniform if it contains at most two bitwise transitions, from 0 to 1 or 1 to 0 (e.g., 00000000, 11000011 and 00000111). For comparison purposes, we have also evaluated the basic LBP, which involves no mapping. The second parameter is the block size, which is used to divide the image into a number of blocks to localize the LBP histograms.

3.2. *WLD*

WLD is a recently developed robust and powerful local descriptor.[27] It consists of two components: differential excitation and gradient orientation. WLD was inspired by a psychological law called "Weber's Law". The law states that "The change of a stimulus (such as sound, lighting) that will be just noticeable is a constant ratio of the original stimulus. When the change is smaller than this constant ratio, a human being would recognize it as background noise rather than a valid signal.[27]" WLD has several advantages, for example, it can reliably extract the edges of an image even in presence of heavy noise, and it is robust against illumination change.

3.2.1. *Differential excitation*

The WLD method computes its first component, the differential excitation, by calculating the ratio between the sum of intensity differences of the center pixel against its neighboring pixels and the intensity of the center pixel. Mathematically, this can be represented as:

$$\xi(x_c) = \arctan\left[\sum_{i=0}^{P-1}\left(\frac{x_i - x_c}{x_c}\right)\right] \tag{4}$$

where x_c is the center/current pixel, $\xi(x_c)$ is the differential excitation of the center pixel, P is the number of neighbors and x_i ($i = 0, 1, \ldots, P-1$) is the ith neighbor of x_c. The arctangent function is used to put limits on the output and avoid the quick increasing or decreasing in the output if the input becomes smaller or larger.[27] The differential excitation can be positive or negative. The positive value indicates that the current pixel is darker than its surroundings and negative values means that the current pixel is lighter than the surroundings. The main purpose of the differential excitation component is to extract the local salient patterns from the image.

3.2.2. *Gradient orientation*

The other component, gradient orientation, is calculated for the center pixel using the following steps[27]:

$$\text{Step1: Compute } \theta = \arctan\left(\frac{V_{10}}{V_{11}}\right) \tag{5}$$

where $V_{10} = x_5 - x_1$, and $V_{11} = x_7 - x_3$. x_5, x_1, x_7, and x_3 are the neighbors of x_c in 3×3 neighborhood as shown below.

x_0	x_1	x_2
x_7	x_c	x_3
x_6	x_5	x_4

$$\text{Step 2: Perform the mapping } f : \theta \to \theta'$$

$$\theta' = \arctan 2\left(v_s^{11}, v_s^{10}\right) + \pi \tag{6}$$

$$\arctan 2\left(v_s^{11}, v_s^{10}\right) = \begin{cases} \theta & v_s^{11} > 0 \text{ and } v_s^{10} > 0 \\ \pi + \theta & v_s^{11} > 0 \text{ and } v_s^{10} < 0 \\ \theta - \pi & v_s^{11} < 0 \text{ and } v_s^{10} < 0 \\ \theta & v_s^{11} < 0 \text{ and } v_s^{10} > 0 \end{cases} \quad \text{where} \quad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \text{ and } \theta' \in [0, 2\pi].$$

$$\text{Step 3: Quantize } \theta' \text{ using the following function:}$$

$$\Phi_t = f_q(\theta') = \frac{2t}{T}\pi, \quad t = \operatorname{mod}\left(\left\lfloor \frac{\theta'}{2\pi / T} + \frac{1}{2} \right\rfloor, T\right) \tag{7}$$

where T is the number of the dominant orientation, and Φ_t is a dominant orientation ($t = 0, 1, \ldots, T-1$). The orientations located inside the interval $\left[\Phi_t - \frac{\pi}{T}, \Phi_t + \frac{\pi}{T}\right]$ are quantized to Φ_t.

3.2.3. *WLD histogram*

We compute WLD histogram from differential excitation and dominant gradient orientation of each pixel. At first, sub-histograms H_t: $t = 0, 1, 2, \ldots, T-1$ of differential excitations corresponding to each dominant orientation Φ_t: $t = 0, 1, 2, \ldots, T-1$ are calculated. All pixels having dominant direction Φ_t contribute to sub-histogram H_t. Then each sub-histogram is further divided into M sub-histograms (segments) $H_{m,t}$: $m = 0, 1, 2, \ldots, M - 1$, each with S bins. Each column of the sub-histogram matrix $H_{m,t}$ corresponds to a dominant direction Φ_t. Each row of this matrix is concatenated as a sub-histogram $H_m = \{H_{m,t}\colon t = 0, 1, 2, \ldots, T - 1\}$. Subsequently, sub-histograms H_m are concatenated into a histogram $H = \{H_\mathrm{m}\colon m = 0, 1, 2, \ldots, M - 1\}$. This histogram represents an image and is referred to as WLD descriptor

The following parameters can affect the performance of WLD:

- T: The number of the dominant orientations used for quantizing the gradient orientation values.
- M: The number of segments used to divide each sub-histogram.
- S: The number of bins in each segment.

In our experiments, we check with different values of T, M, and S to optimize the performance.

3.2.4. *Kruskal-Wallis feature selection*

The number of bins in LBP and WLD histograms is very large, especially when they are calculated in blocks. Many of these bins may not contain sufficient discriminative information, and as a consequence may contribute to lower classification performance. Moreover, dealing with too many features slows down the classification process. One way to deal with this issue is using feature selection. Many types of feature selection techniques have been proposed in the literature.[28,29] Most of them are computationally expensive and complex in nature. In this study, we have adopted the Kruskal-Wallis feature selection method,[29] which is simple to implement and has low computational complexity. The Kruskal-Wallis method is a non-parametric, one-way ANOVA (i.e., analysis of variance) test that can be applied to two or more classes. It tests the hypothesis whether the samples from two or more groups have equal medians, and returns a value p. If p is close to zero for a certain feature, then the feature is selected as it is expected to have good discriminative power. On the other hand, if p is far from zero for some feature, then the feature is expected to have low discriminative value and is discarded.

During training, we apply the Kruskal-Wallis method to LBP and WLD features separately. The features that have p values less than a threshold are selected, and their indices are stored. During testing, we use the selected features only for classification.

3.3. *Classifier*

We have experimented with the minimum distance classifier using three different distance measures: City block distance (L1), Euclidean distance (L2), and Chi-square distance (CS). These distances are defined, respectively, as follows:

$$L1 = \sum_{j=1}^{n} | x_{rj} - x_{sj} | \tag{5}$$

$$L2 = \sqrt{\sum_{j=1}^{n} (x_{rj} - x_{sj})^2} \tag{6}$$

$$CS = \frac{(x_{rj} - x_{sj})^2}{x_{rj} + x_{sj}} \tag{7}$$

where r and s are two images, n is the dimension of the feature vector, and x_{rj} is the jth feature of image r.

3.4. *Fusion of LBP and WLD*

LBP and WLD features do not encode the same information. This is evident from the errors that LBP and WLD produce when tested on the same set of faces. In our experiments, we found that only 15% of the errors are common between LBP and WLD. This indicates that fusing LBP and WLD features might lead to higher classification performance. We have experimented with fusing LBP with WLD features using the following three schemes: (a) feature level fusion by simply concatenating LBP and WLD histograms, (b) score level fusion by averaging LBP and WLD distances, and (c) score level fusion by taking the smallest distance between LBP and WLD distances. Next, we provide more details.

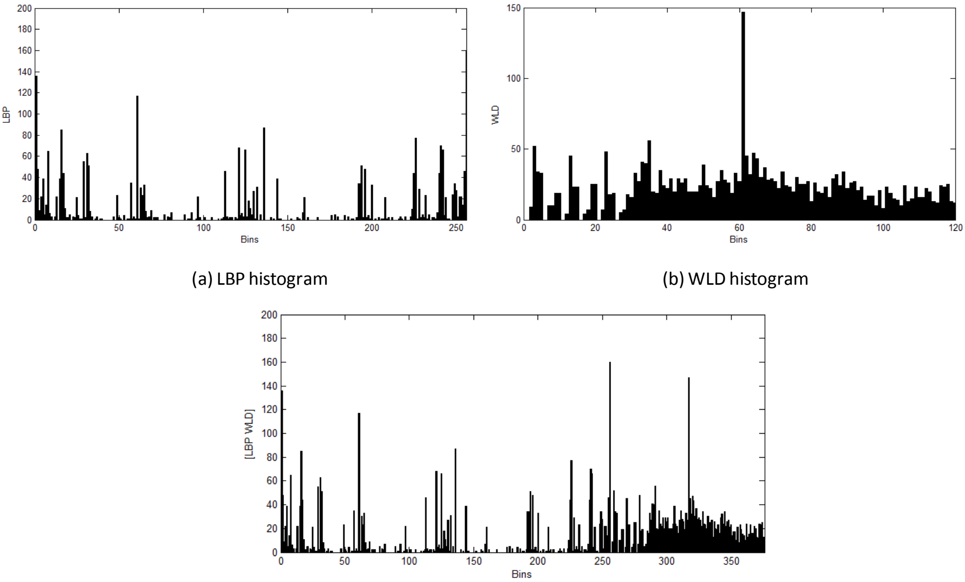
3.4.1. *Feature level fusion*

After optimizing the LBP and WLD histograms using Kruskal-Wallis feature selection, the two histograms are simply concatenated to produce a combined [LBP WLD] histogram. Figure 5 shows an example. Figure 5(a) shows an optimized LBP histogram, 5(b) shows the corresponding optimized WLD histogram, and 5(c) shows the concatenated [LBP WLD] histogram. The minimum distance classifier uses the concatenated histograms for classification.

3.4.2. *Score level fusion: calculating the average*

Score level fusion is performed using the distance matrices of the LBP and WLD histograms. Suppose that L corresponds to one of the three distance measures described in Section 3.4, and L_LBP and L_WLD correspond to the distances obtained using LBP histograms and WLD histograms, respectively. Then, fusion is performed using the following equation:

$$L_i = \frac{\mathrm{L\_LBP}_{test,i} + \mathrm{L\_WLD}_{test,i}}{2} \tag{8}$$



(a) LBP histogram (b) WLD histogram

(c) LBP and WLD concatenated histogram

Fig. 5. Feature level fusion by concatenating (a) LBP histogram and (b) WLD histogram.

where $\mathrm{L\_LBP}_{test,i}$ denotes the distance between a test image and the ith training image using LBP histograms, and $\mathrm{L\_WLD}_{test,i}$ denotes the distance between the same test image and ith training image using WLD histograms. The minimum distance classifier uses L_i for classification.

3.4.3. *Score level fusion: Calculating the minimum*

In this case, score level fusion is performed by taking the minimum distance between $\mathrm{L\_LBP}_{test,i}$ and $\mathrm{L\_WLD}_{test,i}$ instead of the average:

$$L_i = \min\left(\mathrm{L\_LBP}_{test,i}, \mathrm{L\_WLD}_{test,i}\right) \tag{9}$$

The minimum distance classifier uses L_i for matching.

4. Experiments

4.1. *Database*

We used the FERET database[30] in our experiments since most other publically available databases do not possess race information or contain a small number of race groups. The database contains a large number of images acquired during different photo sessions and has a good variety of gender, ethnicity and age groups. The lighting conditions, face orientation and time of capture vary. All faces were normalized in terms of orientation, position and size prior to experimentation. They were also masked to include only the face region (i.e., upper body and background were cropped out) yielding an image size of 60 × 48 pixels. Images were grouped into eight race groups based on the race information provided for each image in the database. The race groups are: White, Asian-Middle-Eastern, Asian, Hispanic, Black-or-African-American, Pacific-Islander, Native-American, and other. Since certain groups contained a small number of images, we used only five major race groups, each containing more than 50 subjects:

- Asian (with 171 subject) – denoted as Asian in the experiments.
- Black-or-African-American (with 78 subject) – denoted as Black in the experiments.
- Hispanic (with 57 subject) – denoted as Hispanic in the experiments.
- Asian-Middle-Eastern (with 53 subject) – denoted as Middle in the experiments.
- White (with 618 subject) – denoted as White in the experiments.

Figure 6 shows several subjects in different race groups.

a. Asian b. Black c. Hispanic d. Middle e. White

Fig. 6. Example faces of different races.

We used 1,188 images for training and 1,180 images for testing. The number of test images per race is given in Table 2.

Table 2. Number of test images per race.

Race Group	Number of Test Images
Asian	190
Black	99
Hispanic	63
Middle	60
White	768

4.2. *Race classification using PCA features*: *Baseline*

We tested race classification using holistic features to establish a baseline for comparisons. For this, PCA features[36] were extracted and compared between training and test images. Three distances were tested as described in Section 3.3. Table 3 shows the classification accuracy (%) obtained using different number of principal components.

As it can be observed, the city block distance performs better than the Euclidean and the chi-square distances. Figure 7 shows a graph of the results for each race in the case of

Table 3. Race classification accuracy (%) using PCA.

Number of Principal Components	Distance	Asian	Black	Hispanic	Middle	White	Average
200	L1	86.84	66.67	65.08	81.67	95.57	79.17
	L2	82.11	70.71	50.79	76.67	94.79	75.01
	CS	84.56	68.89	64.12	80.54	94.37	78.50
300	L1	87.37	66.67	66.67	78.33	95.57	78.92
	L2	81.58	71.72	52.38	76.67	94.66	75.40
	CS	86.84	70.71	62.45	77.33	95.57	78.58
400	L1	89.47	68.69	65.08	78.33	95.96	79.51
	L2	82.11	71.72	52.38	76.67	94.92	75.56
	CS	87.12	71.71	63.75	78.33	95.96	79.37
500	L1	87.89	62.63	66.67	78.33	96.22	78.35
	L2	82.11	71.72	52.38	76.67	94.79	75.53
	CS	85.65	69.33	61.32	76.67	95.35	77.66
600	L1	87.37	62.63	65.08	76.67	96.22	77.59
	L2	82.11	71.72	53.97	76.67	95.05	75.90
	CS	85.65	66.25	60.46	76.67	95.05	76.82

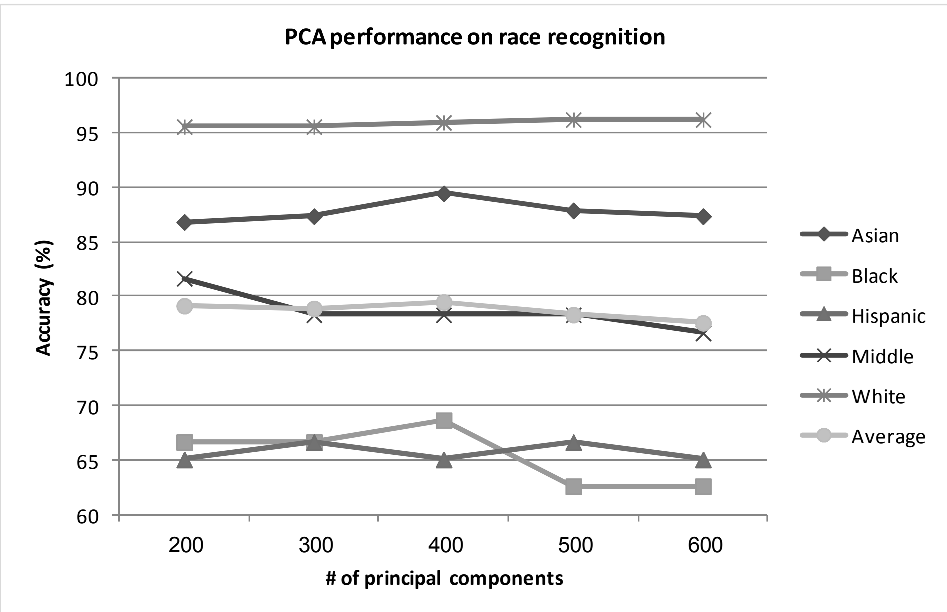


Fig. 7. PCA performance using different number of principal components for race classification. The city block distance was used.

using the city block distance. As it is illustrated, accuracy does not change significantly with the number of principal components. Therefore, we chose 200 principal components for the city block classifier as the baseline for our subsequent experiments.

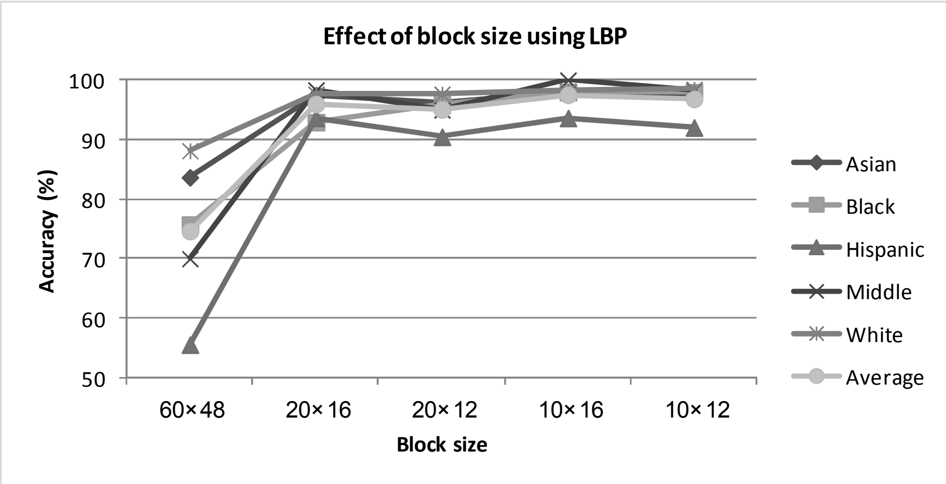
4.3. *Race classification using LBP and WLD features*

The objective of these experiments was to test LBP and WLD separately by optimizing their parameters. For this reason, block size was varied as follows: (20 × 16), (20 × 12), (10 × 16), (10 × 12), and (60 × 48) (i.e., whole image). For the $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{mapping}}$ operator, the following mappings were tested: 'u2' for uniform LBP, 'ri' for rotation-invariant LBP, 'riu2' for uniform rotation-invariant LBP and 0 for no mapping. In separate experiments (i.e., not shown here) it was found that $\mathrm{LBP}_{8,1}^{\mathrm{mapping}}$ worked best; therefore, (P, R) was fixed to (8, 1). The basic LBP method, which involves a 3 × 3 rectangular neighborhood and no mapping, was also used in our experiments.

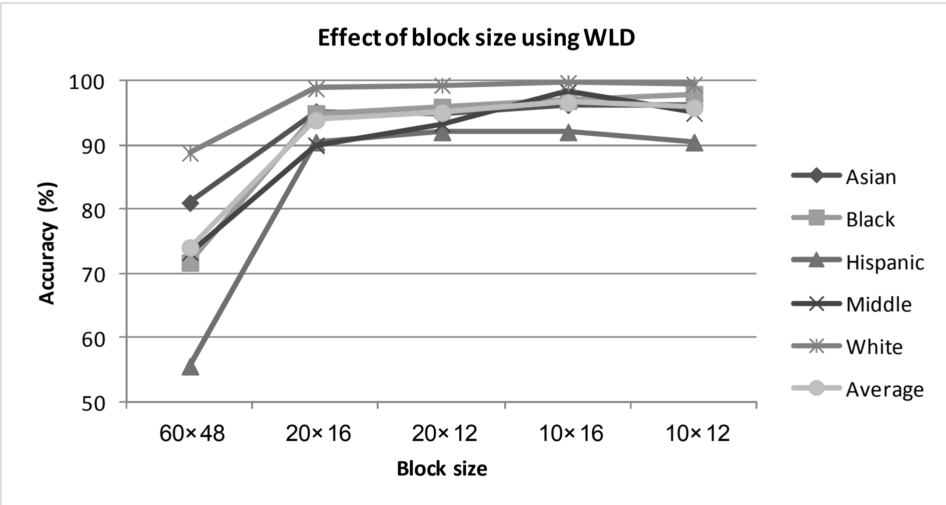
In WLD, we optimized the three parameters that affect its performance: number of dominant orientations (T), number of differential excitation segments (M), and number of bins in sub histogram segments $H_{m,t}$ (S). The influence of these parameters on the method's performance has been discussed in Ref. 27. Specifically, it was found that if these parameters become larger, the dimensionality of the histogram becomes larger; as a result, the histogram becomes more discriminatory. On the other hand, if the parameters become smaller, the histogram becomes statistically more reliable; however, if the parameters become too small, the histogram loses its discriminatory power. The best results reported in Ref. 27 correspond to $M = 6$, $T = 8$ and $S = 20$ on a texture classification application. In our experiments, the values of these parameters were varied as follows: M = 4 or 6; T = 6 or 8; and S = 10 or 15. Combinations of these parameters values were used in our experiments.

Effect of block size: Figure 8 shows the effect of block size on race classification using LBP and WLD features. Our best results, shown in Fig. 8(a), were obtained using basic LPB and the City block distance. In the case of WLD features, our best results were obtained using the City block distance with $T = 8$, $M = 4$, $S = 5$ (see Fig. 8(b)). As it can be noted, using the whole image (60 × 48) has affected performance in all race groups. The average accuracy for LBP and WLD were 74.63% and 74.09% which is worse than that of PCA (i.e., 79.17%). These findings suggest that local descriptors do not perform very well when applied globally.

If LBP or WLD are applied locally (i.e., using smaller blocks), performance increases. The best results in terms of average accuracy were 97.54% and 96.68%, respectively for LBP and WLD, which were achieved using a 10 × 16 block size. However, performance decreases if block size becomes too small; for example, the average accuracy was down to 96.84% for LBP and 95.85% for WLD, using a 10 × 12 block size. This indicates that very small block sizes do not capture sufficient information; therefore, there must be some balance between global and local information to achieve high performance.



(a) Effect block size using LBP; best results (shown here) were obtained using basic LBP and the chi square distance.



(b) Effect of block size using WLD; best results (shown here) were obtained using $[T, M, S] = [8, 4, 5]$.

Fig. 8. Effect of block size on race classification performance using (a) LBP and (b) WLD. Best results were obtained using the city block distance, and (a) Basic LBP and (b) $[T, M, S] = [8, 4, 5]$ for WLD.

Effect of other parameters: Figure 9 shows the effect of LBP mapping parameters on race classification accuracy. These were the best results which were obtained using a 10×16 block size and the City block distance. As illustrated, Basic LBP performs best in all race groups. Rotation invariant (ri) and rotation invariant with uniform mapping (riu2) mappings do no show good results in this particular experiment. In the case of

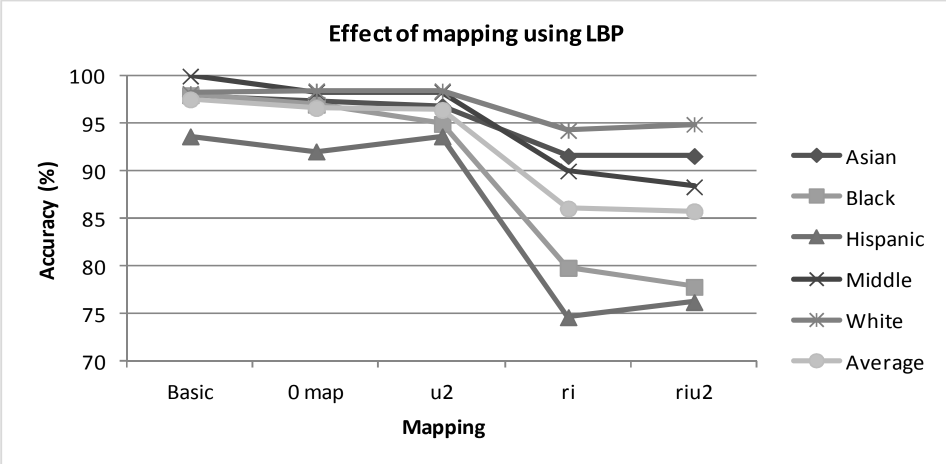


Fig. 9. Effect of LBP mapping parameters on race classification. Best results (shown here) were obtained using a 10 × 16 block size and the city block classifier.

WLD, the parameters T, M, and S do not exhibit any clear trend of decreasing or increasing accuracy, hence we do not provide detailed results in the case of these parameters.

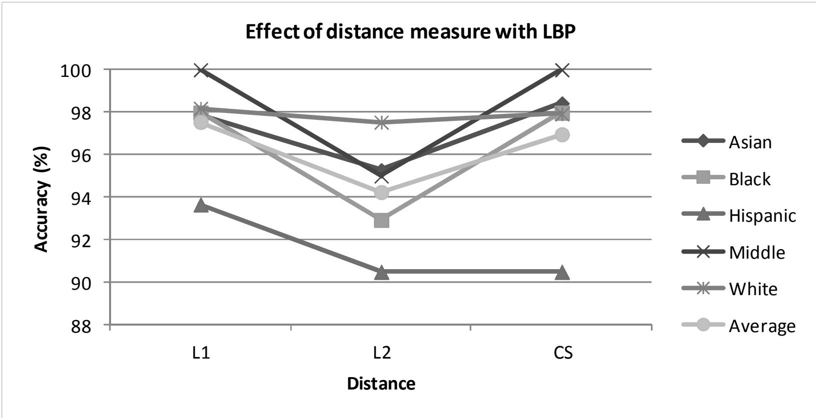
Effect of distance measure: Figure 10 shows the effect of different distance measures on race classification. In most races, the performance of L1 and CS are comparable, while that of L2 is worst. The average accuracy using L1, L2, and CS are 97.54%, 94.24%, and 96.96%, respectively, for LBP, and 96.68%, 94.23%, and 95.87%, respectively, for WLD. Comparing the performance of L1 and CS, L1 performs better than CS.

Table 4 represents the confusion matrix of different race classifications using the best (a) LBP and (b) WLD results. As shown, the confusion of Hispanic faces with White faces is 4.76% using LBP and 6.35% using WLD; these are the highest confused rates in the table. LBP performs better than WLD in all race groups except for White. Figure 11 shows some misrecognized faces using (a) LBP and (b) WLD.

4.4. *Race classification using Kruskal-Wallis feature selection*

The Kruskal-Wallis feature selection technique, described in Section 3, was applied on the best LBP and WLD results to select feature subsets that are important for race classification. As noted in the previous section, the best results were obtained using a 10 × 16 block size and the L1 metric. In the case of LBP, the best results were obtained using basic LBP histogram, where the total number of features was 4608. In the case of WLD, the best result was obtained using [T, M, S] = [8, 4, 5], where the total number of features was 2880. The value of p (significance) was varied to find an optimum threshold for discarding redundant histogram bins. The variation of p was as follows: [0.01 ~ 0.19]

(a)



(b)

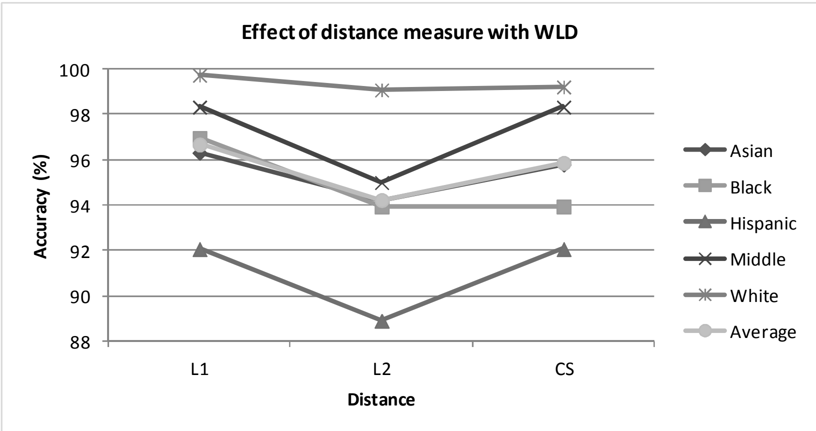


Fig. 10. Effect of distance measures on race classification using (a) LBP and (b) WLD. Best results (shown here) were obtained using a 10 × 16 block size, and (a) the basic LBP and (b) WLD with $[T, M, S] = [8, 4, 5]$.

Table 4. Confusion matrix of different race faces using (a) LBP and (b) WLD.

		Output				
		Asian	Black	Hispanic	Middle	White
Input	Asian	97.89%	–	–	0.53%	1.58%
	Black	–	97.98%	–	–	2.02%
	Hispanic	1.59%	–	93.65%		4.76%
	Middle	–	–	–	100%	–
	White	0.26%	0.13%	0.91%	0.52%	98.18%

(a)

		Output				
		Asian	Black	Hispanic	Middle	White
Input	Asian	96.32%	-	–	1.05%	2.63%
	Black	–	96.97%	1.01%	–	2.02%
	Hispanic	1.59%	–	92.06%	–	6.35%
	Middle	1.67%	–	–	98.33%	–
	White	0.13%	–	0.13%	–	99.74%

(b)

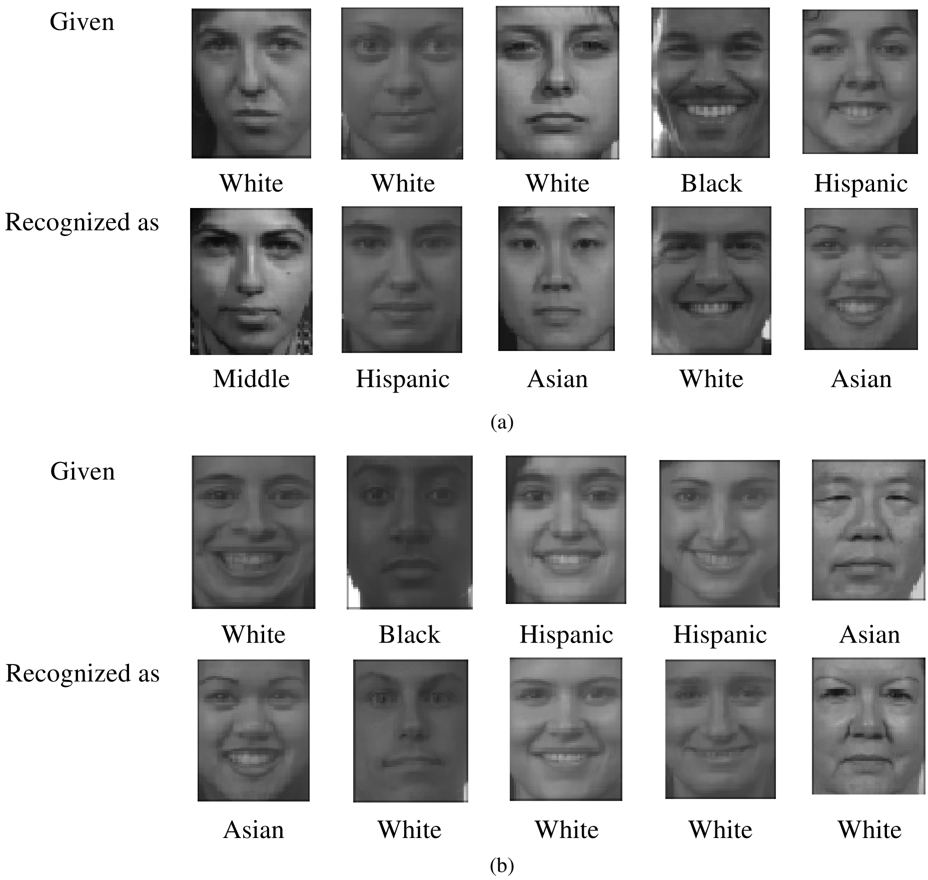
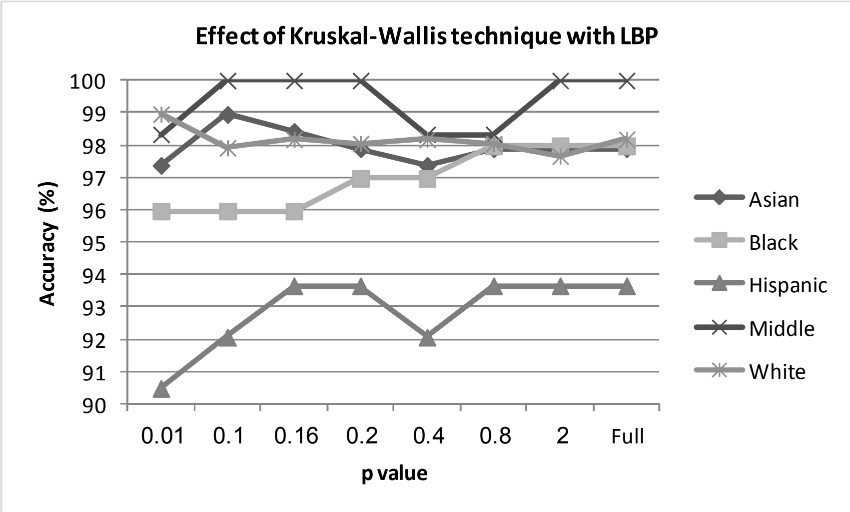
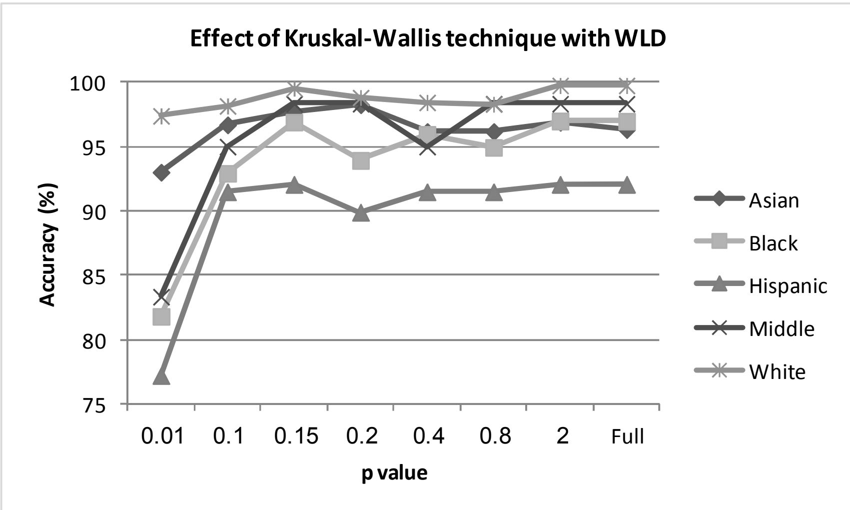


Fig. 11. Erroneous recognized faces using (a) LBP and (b) WLD.

with an increment of 0.03 and [2.0 ~ 4.0] with an increment of 0.2. Figure 12 shows the classification accuracy for each of the five races using feature selection with representative p values. Table 5 shows the number of LBP and WLD features selected for different values of p. Using LBP, very close accuracies were obtained for each race, except for Black, compared to using all histogram bins (i.e., no feature selection). However, only one third of the features was required (i.e., at $p = 0.16$). We obtained similar results using WLD where the number of features was reduced by half (i.e., at $p = 0.16$) without affecting classification accuracy. In subsequent experiments, we use only the subset of bins select at $p = 0.16$. In this case, the accuracies obtained using LBP were: Asian: 98.42%, Black: 95.96%, Hispanic: 93.65%, Middle: 100% and White: 98.18%; similarly, the accuracies obtained using WLD were: Asian: 97.74%, Black: 96.89%, Hispanic: 92.06%, Middle: 98.33%, and White: 99.53%.



(a)



(b)

Fig. 12. Effect of Kruskal-Wallis feature selection using (a) LBP and (b) WLD features. The horizontal axis represents the significance value '*p*'.

Table 5. Number of features selected using different *p* values.

P value →	0.01	0.1	**0.16**	0.2	0.4	0.8	2	Full
# of LBP features	661	1350	**1633**	1795	2558	3968	4598	4608
# of WLD features	752	1433	**1632**	1793	2206	2685	2801	2880

4.5. *Race classification using fusion of LBP and WLD features*

Feature level fusion: To implement feature level fusion, we concatenate the LBP and WLD histogram bins selected by the Kruskal-Wallis (KW) feature selection technique. The total number of bins in the concatenated histogram is (1633 + 1632 = 3265). All the three distance metrics were investigated, and the city block distance (L1) minimum classifier gave the best results.

Score level fusion: Two types of score level fusion were tested: (i) average of LBP and WLD distances, and (ii) minimum of LBP and WLD distances. Three different distance metrics were used to calculate $L\_LBP_{test,i}$ and $L\_WLD_{test,i}$, where chi-square distance (CS) gave the best results.

Table 6 shows the results obtained for each fusion approach along with results obtained without using fusion. Feature level fusion, denoted as Exp. (5)., achieves the highest accuracy in all five race categories; Table 7 shows the confusion matrix for this case. It is worth mentioning that score level fusion, denoted as Exp. (6) and Exp. (7), has higher accuracy compared to LBP or WLD alone. These results suggest that LBP and WLD contain complementary information for face classification and that their fusion improves performance.

Table 6. Results with and without fusion.

Exp.#	Exp. Name	#Features	Accuracy (%)				
			Asian	Black	Hispanic	Middle	White
(1)	LBP	4608	97.89	97.98	93.65	100	98.18
(2)	LBP + feature selection	1633	98.42	95.96	93.65	100	98.18
(3)	WLD	2880	96.32	96.97	92.06	98.33	99.74
(4)	WLD + feature selection	1632	97.74	96.89	92.06	98.33	99.53
(5)	(2) + (4) Feature level fusion	3265	99.47	98.99	96.83	100	100
(6)	(2) + (4) Avg score fusion	3265	99.33	98.32	96.31	100	99.85
(7)	(2) + (4) Min score fusuin	3265	99.33	98.51	96.31	100	99.85

Table 7. Confusion matrix for best performing method (i.e., feature level fusion).

		Output				
		Asian	Black	Hispanic	Middle	White
Input	Asian	99.47%	–	–		1.53%
	Black	–	98.99%	–	–	1.01%
	Hispanic	–	–	96.83%	–	3.17%
	Middle	–	–	–	100%	–
	White	–	–	–	–	100%

4.6. *Elapsed time*

A time comparison between LBP, WLD, and the best performing method based on feature level fusion is shown in Table 8. The time needed for feature selection is only during training. During testing, the methods use the bins selected during training, therefore, they do not have extra time requirements at this stage. The times shown in Table 7 are per face image in seconds. Training time includes the time needed to generate the histogram and to select the bins, while testing time includes the time needed to generate the histogram and to find the distances between a test image histogram and all training image histograms to find the minimum distance. As shown in the table, the best performing method takes 0.7073 seconds for training and 0.3337 seconds for testing. This amounts to almost the sum of time needed for LBP and WLD.

Table 8. Elapsed time in seconds for different methods.

Methods	Training (Histogram Generation and/or Selection)	Testing (Histogram Generation and Classification)
LBP	0.1363	0.1394
LBP + feature selection	0.1597	0.1378
WLD	0.1952	0.1979
WLD + feature selection	0.2161	0.1968
LBP + WLD feature level fusion	0.7073	0.3337

5. Conclusion

We have investigated the performance of LBP and WLD features for race classification. The Kruskal-Wallis feature selection algorithm was used to select an optimal subset of bins from the LBP and WLD histograms. We experimented with different fusion schemes to improve classification accuracy. The best performing method was obtained by concatenating the reduced size histograms. The following conclusions can be drawn based on our experimental results:

(a) Feature level fusion outperforms PCA, LBP, and WLD in all race groups in terms of accuracy. Feature level fusion achieves 100% accuracy for the Middle and White races. The lowest accuracy (i.e., 96.83%) was obtained for the Hispanic race; however this was the best accuracy for this race group among all other methods.
(b) Using local features (i.e., using LBP and WLD with small windows) yielded higher classification accuracy than using holistic features (i.e., PCA features or LBP/WLD features based on large windows).
(c) LBP performed better than WLD in all race groups except White.
(d) Using the Kruskal-Wallis feature selection method, comparable accuracy was obtained using one-third of the features, in the case of LBP, and one-half of the features, in case of WLD.
(e) Both feature level fusion and score level fusion using LBP and WLD gave higher classification accuracy than using LBP and WLD alone.

For future work, we plan to investigate feature level fusion of LBP and WLD features for improving the performance of gender and age classification. In this context, we plan to experiment with more types of local features and feature selection algorithms.

Acknowledgments

This work is supported by the National Plan for Science and Technology, King Saud University, Riyadh, Saudi Arabia under project number 10-INF1044-02. George Bebis is a Visiting Professor in the College of Computer and Information Sciences at King Saud University, Riyadh 11543, Saudi Arabia.

References

1. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, Face recognition: A literature survey, *ACM Computing Surveys* **35**(4) (2003) 399–458.
2. A. Abate, N. Nappi, D. Riccio, and G. Sabatino, 2D and 3D face recognition: A survey, *Pattern Recognition Letters* **28**(14) (2007) 1885–1906.
3. J. Brigham, The influence of race on face recognition, in *Aspects of Face Processing*, eds. H. Ellis, M. Jeeves, and F. Newcombe (1986), pp. 170–177.
4. A. O'Toole, J. Peterson, and K. Deffenbacher, Another-race effect for classifying faces by sex, *Perception* **25** (1996) 669–676.
5. Y. Cheng, A. O'Toole, and H. Abdi, Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding, *Cognitive Science* **25** (2001).
6. M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* **3** (1991) 71–86.
7. G. Guo and 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* (*CVPRW*), (June 2010), pp. 79–86.
8. T. Ahonen, A. Hadid, and M. Pietikäinen, Face description with local binary patterns: Application to face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28** (2006) 2037–2041.
9. T. Ojala, M. Pietkainen, and D. Harwood, A comparative study of texture measures with classification based on feature distributions, *Pattern Recognition* **29** (Jan. 1996) 51–59.
10. C. Liu and H. Wechsler, A gabor feature classifier for face recognition, in *Proc. 8th IEEE Int. Conf. Computer Vision* (*ICCV*), (2001), pp. 270–275.
11. D. G. Lowe, Distinctive image features from scale invariant keypoints, *International Journal of Computer Vision* **60** (Nov. 2004) 91–110.
12. C. Geng and X. Jiang, SIFT feature for face recognition, in *Proc. 2nd IEEE Int. Conf. Computer Science and Information Technology*, (2009), pp. 553–698.
13. R. Brunelli and T. Poggio, Face recognition: Features versus templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(10) (1993) 1042–1052.
14. N. Dalal, and B. Triggs, Histogram of oriented gradient for human detection, in *Proc. Computer Vision and Pattern Recognition* (*CVPR*), (2005), pp. 886–893.
15. B. Zhang, Y. Gao, S. Zhao, and J. Liu, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, *IEEE Transactions On Image Processing* **19**(2) (February 2010) 533–544.
16. G. Cottrell and M. Fleming, Face recognition using unsupervised feature extraction, *International Conference on Neural Network*, (1990), pp. 322–325.

17. L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, Face recognition and gender determination, *International Workshop on Automatic Face and Gesture Recognition,* (1995), pp. 92–97.
18. J. Fellous, Gender discrimination and predication on the basis of facial metric information, *Vision Research* **37** (1997) 1961–1973.
19. A. Jain, R. Duin, and J. Mao, Statistical pattern recognition: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1) (2000) 4–37.
20. O. H. MacLin and R. S. Malpass, Racial categorization of faces: The ambiguous race face effect, *Psychology, Public Policy, and Law* **7**(1) (2001) 98–118.
21. P. J. Phillips, F. Jiang, A. Narvekar, J. Ayyad, and A. O'Toole, An other-race effect for face recognition algorithms, *ACM Transactions on Applied Perception* **8**(2) (2011).
22. X. Lu and A. K. Jain, Ethnicity identification from face images, *Proc. of SPIE International Symposium on Defense and Security: Biometric Technology for Human Identification*, (2004), pp. 114–123.
23. S. Hosoi, E. Takikawa, and M. Kawade, Ethnicity estimation with facial images, in *Proc. 6th IEEE Int. Conf, on Automatic Face and Gesture Recognition* (*AFGR*), (2004), pp. 195–200.
24. X. Lu, H. Chen, and A. K. Jain, Multimodal facial gender and ethnicity identification, in *Proc. International Conf. on Biometrics* (*ICB*), (Hong Kong, 2006), pp. 554–561.
25. Y. Zhiguang and A. Haizhou, Demographic classification with local binary patterns, in *Proc. International Conf. on Biometrics* (*ICB*), (2007), pp. 464–473.
26. F.S. Manesh, M. Ghahramani, and Y.P. Tan, Facial part displacement effect on template-based gender and ethnicity classification, *11th International Conference on Control Automation Robotics & Vision* (*ICARCV*), (2010), pp. 1644–1649.
27. J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, WLD: A robust local image descriptor, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9) (Sep. 2010) 1705–1720.
28. L. Huan, M. Hiroshi, S. Rudy, and Z. Zheng, Feature selection: An ever evolving frontier in data mining, in *Proc. The Fourth Workshop on Feature Selection in Data Mining* (*FSDM10*), (2010), pp. 4–13.
29. S. Yvan, I. Inaki, and L. Pedro, A review of feature selection techniques in bioinformatics, *Bioinformatics* **23**(19) (2007) 2507–2517.
30. P. J. Phillips, M. Hyeonjoon, S. A. Rizvi, and P. J. Rauss, The FERET evaluation methodology for face-recognition algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10) (Oct. 2000) 1090–1104.
31. S. Gutta, H. Wechsler, and P. J. Phillips, Gender and ethnic classification of face images, in *Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998. *Proceedings*, (1998), pp. 194–199.
32. U. Tariq, Y. Hu, and T. S. Huang, Gender and ethnicity identification from silhouetted face profiles, in *16th IEEE International Conference on Image Processing* (*ICIP*), (2009), pp. 2441–2444.
33. G. Zhang and Y. Wang, Multimodal 2D and 3D facial ethnicity classification, in *Fifth International Conference on Image and Graphics*, 2009 (*ICIG* '09), (2009), pp. 928–932.
34. X.-dong Duan, C.-rui Wang, X.-dong Liu, Z.-jie Li, J. Wu, and H.-long Zhang, Ethnic Features extraction and recognition of human faces, (2010), pp. 125–130.
35. G. Shakhnarovich, P. A. Viola, and B. Moghaddam, A unified learning framework for real time face detection and classification, in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002. *Proceedings*, (2002), pp. 14–21.
36. M. Kirby and L. Sirovich, Application of Karhunen-loeve procedure for the characterization of human faces, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(1) (Jan. 1990) 103–108.