Face recognition using Weber local descriptors

Shutao Li, Dayi Gong, Yuan Yuan

Abstract

This paper presents a method for face recognition using multi-scale Weber local descriptors (WLDs) and multi-level information fusion. Our method introduces the WLD, a novel and robust local descriptor, to describe the facial images and modifies it by a non-linear quantization approach to enhance its discriminative power. Moreover, a multi-scale framework for WLD extraction with multi-level information fusion approaches is provided for face representation and recognition. The proposed method has four main steps: (1) image partition: under given rules, each facial image is uniformly divided into a set of non-overlapped sub-regions; in this way, for a set of facial images, we therefore have a large pool of this type of sub-regions; (2) feature extraction: in this pool of sub-regions, taking one sub-region as a center, a group of similar ones are chosen for extraction of WLD histogram features; (3) features measurement: these WLD histograms are then fused into a single vector – as the feature of the center sub-region. Nearest neighborhood on chi-square is employed for similarity measurement between two sub-regions; and (4) voting: the recognition result of the entire probe (a face in sub-regions) is obtained via a voting function on the recognition result of all its sub-regions. Experimental results demonstrate the effectiveness of the proposed method upon three popular datasets.

Keywords:
Face recognition
Facial image analysis
Weber local descriptors (WLDs)
Information fusion

1. Introduction

Face recognition has been standing as a popular task, because of its wide range of applications. With the developments of image processing, pattern recognition and computer technology, a lot of face recognition methods have been proposed and achieved good performance under the controlled conditions [1]. Generally, these face recognition methods could be summarized into two categories: holistic matching methods and local matching methods [2].

The holistic matching methods take a single feature, which is extracted from the whole face image, as input for face recognition. Most of them project the facial images into some low dimensional subspaces or sub-manifolds that preserve certain intrinsic properties [3]. Some works, such as principal component analysis (PCA) [4–7], independent component analysis (ICA) [8–10], focus on finding the subspaces preserving certain distributive properties; Some methods, including linear discriminant analysis (LDA) [11,12] and maximization of the geometric mean of all divergences (MGMD) [13], attempt to find a transformation to project facial images into the subspaces that preserve certain discriminative property: Efforts are made to find the subspaces to preserve some locality properties in some subspace technologies, such as isometric feature mapping (ISOMAP) [14,15], locality preserving projection (LPP) [16,17], local linear embedding (LLE) [18], and other manifold learning methods [19,20]. Most traditional subspace selection methods assume that the data obeys the Gaussian distribution, which satisfies the case that the data just suffers from the change of one factor. While the tensor subspace analysis considers the facial image as a linear combination of multiple factors, such as identity, expressions, poses illumination conditions, and provides ways of analyzing their influences [21–24]. Compared to these methods mentioned above, nonnegative matrix factorization (NMF) projects the facial image into the subspaces with non-negativity constraints [25–27]. The performance of the holistic matching methods obviously degrades with variations in expressions, poses, illumination conditions, partial occlusions, etc. Increasing attention recently has been put into the local matching methods.

In contrast to the holistic matching methods, the local matching methods are more robust in uncontrolled conditions. They generally divide facial images into several components, the features of which are extracted respectively. Then the information fusion methods are generally adopted to combine the feature of each component into a single feature for further recognition or to combine the recognition result of each component. The local binary pattern (LBP) [28,29] and Gabor wavelet feature [30–32] are two of the most popular local features for face recognition. The LBP describes the distribution of some local micro-patterns of images, such as edges, spots and flats; while the Gabor wavelet...
feature can be considered as a detector about orientation and scale tunable edge, etc. They have been widely used in image analysis, such as texture classification and segmentation, image registration and face recognition. Some local features by combining of LBP and Gabor feature have been proposed [33–36], which can achieve better recognition performance than either feature alone. The LBP is performed on Gabor magnitude information (LGBP-M) [34] and Gabor phase information (LGBP-P) [35]. In [36], the LBP is adopted to describe the neighboring relationship not only in image space, but also in the responses by convolving multi-scale and multi-orientation Gabor filters.

The local matching methods have shown promising recognition performance, but there still exist challenges that the face recognition task has to confront. These challenges can be summarized in two factors: one is about the large variations in face images of the same person in different environments; and the other is the contradictory between high dimensionality of face data and small sample size [1]. Aiming at dealing with these challenges, we propose a method for face recognition using multi-scale Weber local descriptors (WLDs) [37] and multi-level information fusion in this paper.

The WLD is derived from Weber’s Law, which states that only if the ratio of the change of a stimulus to original stimulus is enough big, this change can be recognized. According to Weber’s Law, two ratios are introduced to describe the texture features of images to obtain its two components: differential excitation and orientation. The WLD first computes the salient micro-patterns by differential excitation, and then builds statistics on these salient patterns along with the gradient orientation of current pixel. Both the WLD and LBP are dense descriptors computed for every pixel and along with the gradient orientation of current pixel. Both the WLD, a powerful and robust local descriptor, to represent facial images for better performance. In our work, we introduce multi-scale sub-images can be conducive to extract more discriminative and robust features of different facial local structures and be seen to increase the number of training samples. In addition, only a part of facial images’ qualities degrade due to variations in facial expressions, illumination conditions and occlusions. We employ a decision fusion rule to reduce the negative influence of local regions with poor quality.

The rest of this paper is organized as follows. Section 2 describes the WLD feature in details. The face recognition approach based on WLD of multi-scale sub-images and multi-level information fusion is proposed in Section 3. Experimental results and conclusions are presented in Sections 4 and 5 respectively.

2. Weber local descriptor (WLD)

In this section, the basic calculation method of WLD is first described. And then we propose a calculation method of WLD for face recognition according to the characteristics of facial images.

2.1. Basic WLD

The WLD uses two ratios to calculate its two components: differential excitation and orientation. Then a concatenated 2D histogram about them is constructed to represent image.

2.1.1. Differential excitation

Differential excitation comes from the ratio between two terms: one is the sum of differences of a current pixel against its neighbors; the other is the current pixel itself. We can calculate it as follows:

\[ \mathbf{v}_1 = \text{lof}_{f_1}, \]

\[ \mathbf{v}_2 = \text{lof}_{f_2}, \]

where \( I \) is the input image, \( \text{lof} \) represents the convolution, \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) are the outputs of filters \( f_1 \) and \( f_2 \) respectively, as shown in Fig. 1. The ratio \( G_1 \) of \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) is mapped into \( [-\pi/2, \pi/2] \) by the arctangent function:

\[ \alpha = \arctan(G_1) = \arctan(\mathbf{v}_1/\mathbf{v}_2). \]

Then \( \alpha \) is linearly quantized into \( T_1 \) dominant differential excitations as follows:

\[ \{\xi_1\} = \text{floor}(\frac{\alpha + \pi/2}{\pi/T_1}), \quad i = 0, 1, 2, ..., T_1-1, \]

where floor(x) is a function, which returns the largest integer less than or equal to x. The differential excitations \( \alpha \) within \( [(-1/2)\pi + (1/2/T_1)\pi, (-1/2)\pi + (1/2/T_1)\pi] \) are consequently quantized to \( \xi_1 \).

2.1.2. Orientation

The orientation component is about the ratio of the change in horizontal direction to that in vertical direction of current pixel. It is actually the gradient orientation obtained using Sobel operator,

\[ \frac{I_x}{I_y} = \frac{f_1}{f_2} \]

\[ \frac{I_y}{I_x} = \frac{f_3}{f_4} \]

Fig. 1. Filter windows of WLD. The filters \( f_1 \) and \( f_2 \) are used to calculate the differential excitation. The filters \( f_3 \) and \( f_4 \) are used for the calculation of gradient orientation.
which is calculated by
\[ \theta = \arctan(G_2) = \arctan \left( \frac{v_3}{v_4} \right) \tag{4} \]

where \(v_3\) and \(v_4\) are the outputs of the filters \(f_3\) and \(f_4\), which are shown in Fig. 1. \(\theta\) is located within \([-\pi/2, \pi/2]\). According to the values of \(v_3\) and \(v_4\), \(\theta\) can be mapped to \(\theta' \in [0, 2\pi]\) by
\[ \theta'(x, y) = \begin{cases} 0 & v_3(x, y) > 0, v_4(x, y) > 0, \\ \theta(x, y) + \pi & v_3(x, y) < 0, v_4(x, y) > 0, \\ \theta(x, y) + \pi & v_3(x, y) < 0, v_4(x, y) < 0, \\ \theta(x, y) + 2\pi & v_3(x, y) > 0, v_4(x, y) < 0. \end{cases} \tag{5} \]

\(\theta'\) is further linearly quantized into \(T_2\) dominant orientations as follow:
\[ \{\psi_j\} = \text{floor} \left( \frac{\theta' - \pi}{2\pi/T_2} \right), \quad j = 0, 1, 2, ..., T_2 - 1. \tag{6} \]

So the orientations within \((2j - 1)/T_2, (2j)/T_2)\) are quantized into \(\psi_j\).

### 2.1.3. WLD histogram

The 2D concatenated histogram \(\{\text{WLD}(\xi, \psi_j)\}\) about the differential excitation and orientation can be constructed to represent image. As is shown in Fig. 2, each row of 2D WLD histogram corresponds to a dominant differential excitation \(\xi_i\), and each column corresponds to a dominant orientation \(\psi_j\). Thus, in this 2D histogram, the intensity of each cell denotes the frequencies of a certain dominant differential excitation on a certain dominant orientation. To obtain a simpler descriptor, the 2D WLD histogram is encoded into 1D histogram \(H\) with the length of \(T_1 \times T_2\). Algorithm 1 summarizes the calculation procedure of basic WLD.

**Algorithm 1.** Basic WLD.

1: **Input:** Image \(I\).
2: Compute its differential excitation \(\alpha\) by (2), \(\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]\).
3: Compute the linearly quantified differential excitation \(\xi_i\) by (3).
4: Compute its gradient orientation \(\theta\) by (4), \(\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]\).
5: Map \(\theta\) to \(\theta'\) by (5), \(\theta' \in [0, 2\pi]\).
6: Compute the linearly quantified orientation \(\psi_j\) by (6).
7: Construct the 2D concatenated histogram \(\{\text{WLD}(\xi_i, \psi_j)\}\) and then encode it into a 1D feature vector \(H\).
8: **Output:** the WLD feature vector \(H\).

### 2.2. Modified WLD for face recognition

In this subsection, distribution characteristics of differential excitations and orientations of face images are analyzed. The average histograms of differential excitations and orientations on

\[ \text{Fig. 2. The 2D WLD histogram. The intensity of each cell corresponds to the frequencies of a certain dominant differential excitation and a certain dominant orientation.} \]

seen that the majority of differential excitations of face image are located on the range of \([-\pi/4, \pi/4]\), especially near peak point, where the value of differential excitation equals to zero. If we use the linear quantization as mentioned above, most pixels in face images would be taken as the same dominant differential excitations near the peak and their distinctions would be overlooked. Similarly, there are four peaks in the average histogram of orientations of facial images at \(0\) \((2\pi)\), \(\pi/2\), \(\pi\), \(3\pi/2\). In our
implementation, we use a non-linear quantization to calculate the differential excitations $\phi_i$ and orientations $\psi_j$ by narrowing the ranges of dominant ones near the peak and widening the ranges where the frequency is small. The specific quantization rules are presented in Section 4. Algorithm 2 summarizes the calculation procedure of WLD for face recognition. We non-linearly quantify the differential excitation and orientation according to their distribution characteristics.

3. Our face recognition method

In this section, we describe the face recognition based on the WLD of multi-scale sub-images and multi-level fusion in details. Our method consists of three basic stages: (1) face images preprocessing, (2) face representation based on the WLD of multi-scale sub-images and (3) recognition based on decision fusion, as shown in Fig. 5.

3.1. Preprocessing

In order to smooth the noises from the capture and the influence of size normalization of face images, the Gaussian filter is adopted:

$$I' = I * G(x, y, \delta),$$

where $*$ represents the convolution, $G(x, y, \delta)$ is obtained by

$$G(x, y, \delta) = \frac{1}{\sqrt{2\pi}\delta} \exp\left(-\frac{x^2 + y^2}{2\delta^2}\right),$$

where $\delta$ is the standard deviation of Gaussian function.

3.2. Feature extraction

In this subsection, a method for face representation using WLD within a multi-scale framework is proposed. The traditional local matching methods for face recognition generally divide the whole facial image into non-overlapped or overlapped sub-images with a fixed size to extract feature respectively. However, different facial local structures (eyes, nose, mouth and skin textures) are shown in different scales image patches. In addition, the locations of local structures will vary with significant facial expression and pose variations. It is not appropriate to divide the facial image into a set of fixed scale sub-images to describe different local structures. As is shown in Fig. 6, texture information that sub-images with different sizes include is different. In our method, we extract features from a set of multi-scale sub-images at the same location in facial images and then fuse them into an enhanced feature. The face representation based on local features of multi-scale sub-images is benefit to extract more discriminative features of different facial local structures. Moreover, it can be considered to increase the samples for recognition to improve the performance.

The method for face representation is described in details as follows. The differential excitations $\phi_i$ ($i=1,2,\ldots,T_1$) and orientations $\psi_j$ ($j=1,2,\ldots,T_2$) of the whole face image are first calculated as in Algorithm 2. The differential excitation image and orientation image are as shown in Fig. 7. And then the facial images are uniformly divided into $N(k \times k)$ rectangular sub-images $S_{n0}$ ($n=1,2,\ldots,N$). As is shown in Fig. 6, most texture features of face (eyes, nose and mouth) are not located in the edge of facial images. So each internal sub-image $S_{n0}$ obtained is taken as center to get $M$ sub-images $S_{nm}$ ($m=1,2,\ldots,M$) with increasing sizes. The corresponding sub-regions $S_{nm}$ of differential excitations $\phi_i$ and the corresponding sub-images $S_{nm}$ of orientations $\psi_j$ are also extracted independently. According to $S_{nm}$ and $S_{nm'}$, a 2D concatenated histogram $W_{nm} = \{WLD(\phi_i, \psi_j)\}$ of the sub-image $S_{nm}$ is calculated. Each column of 2D histogram $W_{nm}$ is used to form a 1D WLD

---

**Algorithm 3.** Face representation using WLD in a multi-scale framework.

1: **Input:** Face image $I$.
2: Uniformly divide the image $I$ into $N$ non-overlapped sub-images $S_{n0}$ ($n=1,2,\ldots,N$).
3: **for** each non-overlapped sub-image $S_{n0}$
   **if** $S_{n0}$ is an internal sub-image
      Taking $S_{n0}$ as the center, extract $M$ sub-images $S_{nm}$ ($m=1,2,\ldots,M$) with growing sizes.
      **for** each sub-image $S_{nm}$ ($m=0,1,\ldots,M$) in a group
         Extract its WLD feature $H_{nm}$ as Algorithm 2.
      **end**
      Combine $H_{nm}$ into an enhanced feature $H_{n}$
   **else**
      Extract its WLD feature $H_{n0}$ as Algorithm 2.
      $H_{n} = H_{n0}$.
   **end**
   **end**
4: **Output:** $N$ feature vectors $H_{n}$ ($n=1,2,\ldots,N$).
histogram $H_{nm}$ in sequence. The enhanced histogram $H_n$ is obtained by combining $M+1$ histograms $H_{nm}$ based on the same location. As a result, we can yield $N$ feature vectors $H_n$ with the length of $K \times T_1 \times T_2$ ($K=1$ or $M+1$). Since internal ones $H_n$ are enhanced by combining histograms of sub-images with different sizes at the same location, their $K$ equal to $M+1$. While $H_n$ are based on the sub-images on the edge of face image, their $K$ equal to $1$. In Algorithm 3, we summarize the procedure of face representation using WLD within a multi-scale framework.

3.3. Matching and decision fusion

The matching results of feature vectors of the test face image are calculated by nearest neighborhood method based on Chi-square distance, and then the final recognition result is obtained by decision fusion.

The Chi-square was used to measure the similarity of two histograms and achieved good performance [29]. Motivated by the successful application, we adopt the Chi-square distance to measure similarity of two WLD histograms of different images. It is defined as

$$\chi^2(x, y) = \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{x_i + y_i},$$  \hspace{1cm} (9)$$

where $\chi^2(x, y)$ represents the Chi-square distance of two vectors $x$ and $y$, and $x_i$ and $y_i$ are their $i$-th bins. As is shown in (9), the Chi-square distance employs ratios of difference and sum of the corresponding bins $x_i$ and $y_i$, while the Euclidean distance only takes the difference of corresponding bins into consideration. Therefore, the Chi-square distance also agrees with the idea of the Weber’s Law as WLD. In our case, the changes near peak of two WLD histograms of different face images are much greater than that of other regions. If we use the Euclidean distance to measure the difference of the WLD histograms of two face images, the regions near the peak will stand out, while the differences of the more regions will be ignored.

The decision fusion is used to combine the matching result of each feature vector of test face image. The quality of local regions of face images will degrade because of the variations in facial expression, illumination and occlusions. Decision fusion can be considered as making a decision by combining the results obtained in different ways. In our case, the matching results based on different locations are combined by voting to obtain the result of the whole face image. In this way, the impacts of the poor quality regions in face image on recognition performance can be reduced.

The chi-square distances $(\chi^2)^d_{nt}$ between each of histograms of feature vectors $H_n^d$ of probe image based on different locations and counterparts $H^d_c$ of training images are calculated by

$$d_{nt} = \sum_{i=1}^{K \times T_1 \times T_2} \frac{((H_n^d(t) - H^d_c(t))^2)}{H^d_c(t) + H^d_c(t)}, \quad d=1, 2, \ldots, D,$$

where $H_n^d$ represents $n$-th histogram vector of $d$-th gallery image. $D$ represents the number of the gallery images. Matching result $r_n$ of each vector is obtained by the minimal rule:

$$r_n^d = \min_d (\chi^2)^d_{nt}. \hspace{1cm} (11)$$

The matching result $r_n$ of $H_n^d$ is the class, which the gallery image corresponding to the minimal Chi-square distance $\chi^2_{nt}$ belongs to.

The recognition result $R$ of the whole probe image is obtained by voting according to each matching result $r_n$ by

$$R =\arg\max_{c=1}^{C} \left(\sum_{n=1}^{N} \mu^c_n \right), \quad \mu^c_n = \begin{cases} 1, & r_n \in \omega_c, \\ 0, & r_n \notin \omega_c, \end{cases} \hspace{1cm} (12)$$

where $C$ is the number of different persons in gallery images and $\omega_c$ represents $c$-th class. $R$ is the class corresponding to the most votes. In Algorithm 4, we summarize our method for face recognition.

**Algorithm 4.** Our method for face recognition.

1: \textbf{Input:} Training face images $I_i$ ($d=1, 2, \ldots, D$), Test face image $I_0$, 
2: Smooth each face image by (7), 
3: Extract the feature vector $H^d_i$ of each training face image $I_i$ and the feature vector $H^d_0$ of test face image $I_0$ ($n=1, 2, \ldots, N$) by Algorithm 3,
4: for each feature vector $H^d_0$ of test face image $I_0$
Compute the Chi-square distance \(\chi^2_{n}\) of corresponding vector \(H^d_n\) of each training image \(I_d\) by (10).

5: Compute the matching result \(r_n\) of each vector \(H^o_n\) of test image \(I_0\) by nearest neighborhood method.
6: The recognition result \(R\) of the whole test image is obtained by (12).
7: Output: The recognition result \(R\).

4. Experiments

The performance of our method for face recognition is demonstrated experimentally on three well-known face image databases, namely, the Yale, AR and FERET databases.

The Yale face database contains 165 facial images of 15 individuals. Each person has 11 different images captured under various illumination conditions (left-light, center-light and right-light), facial details (with glasses or without) and facial expressions as shown in Fig. 8. Each face image in this database is cropped and resized to 100 × 100.

The AR face database [39] appears challenging to many outstanding face recognition algorithms due to the large number of individuals and total face images. The face images of the AR face database reach the number of 4000, grabbed from 126 people (70 men and 56 women). In our case, the subsets (50 men and 50 women) [40] are chosen to conduct experiments, which are taken in two sessions separated by two weeks. Each session contains 13 images per person, including one neutral face, three different expressions such as smile, anger and screaming, three different illumination conditions (left light on, right one on and all side lights on), three different wearing sunglasses under different illuminations and three wearing scarf under different illuminations, as shown in Fig. 9. Each image is cropped and resized to 100 × 100 in our experiments.

The FERET face database [41] which has been available for testing and evaluating state-of-the-art face recognition algorithms

Fig. 8. Face images of one person on Yale database.

Fig. 9. Face images of one person on AR database, face images: (a) and (b) are taken in two sessions separated by two weeks.
is obtained by the FERET face program sponsored by the US Department of Defense through the DARPA Program, with 14,126 images from 1199 individuals included. In our experiments, we select a subset of the FERET database containing 1400 images of 200 individuals (each individual has seven images whose names are marked with two-character string: “ba”, “bd”, “be”, “bf”, “bg”, “bj”, and “bk.”) [42]. Images of the subset differ from each other in facial expression, illumination and pose (±15° and ±25°), as shown in Fig. 10. Each image in this database is cropped and resized to 80×80 in our experiments.

4.1. Experiments setting

4.1.1. The rule of non-linear quantization

There are two similar parts in the average histogram of differential excitations and each of them is non-linearly quantized into four dominant differential excitations, as shown in Fig. 11. The non-linear quantization rule of differential excitations by integrating unequal ranges of the histogram is presented in Table 1 in details.

As is shown in Fig. 12, there exist four similar parts in the histogram of orientations and each of them is non-linearly quantized into four dominant orientations. The quantization rule of orientations by integrating unequal ranges of the histogram is presented in Table 2.

4.1.2. Discussion on parameters

Experiments are conducted on AR database and FERET database to explore the effect on recognition performance of parameters N and M. N is the number of non-overlapped sub-images and M is the number of sub-images with different sizes based on the obtained non-overlapped sub-images. We use the two neutral images per person in the two sessions of AR database as the training set, leaving the four subsets (facial expression, illumination, occlusion by wearing sunglasses or scarf) as testing sets respectively, which is similar to [36]. On FERET database, the first l (l=2–6) images per person are used for training and the remaining for testing, as the same with in [42]. At first, we uniformly divide the face image into N (k×k) non-overlapped sub-images and do not extract any more sub-image based on the same location (M=0). k represents the number of sub-images divided per row or per column. We serve N (k×k) as 9 (3×3), 16 (4×4), 25 (5×5), 36 (6×6), 49 (7×7), 64 (8×8), 81 (9×9), 100 (10×10) in sequence to discover its impact on recognition performance of our method. When the parameter δ of Gaussian filer is 0.8, we can obtain good performance. In all following experiments, δ is set to 0.8.

The results of the experiments on parameter N (k×k) on AR database are shown in Fig. 13, where the horizontal axis denotes the number N of non-overlapped sub-images and the vertical axis describes the recognition rate. The results show that WLD is strongly robust to facial expression since the corresponding recognition rate maintains above 90% with the facial expression (±15° and ±25°) growing. Especially for the scarf probe set, the recognition rate maintains above 90% with the facial expression (±15° and ±25°) growing. The recognition rate rapidly as N grows. Especially for the scarf probe set, the
recognition rate soars from 60.67% to 91.17% when $N$ ranges from 9 (3 × 3) to 81 (9 × 9). As a consequence of the variations in illumination and occlusion, the quality of a part of face image decreases. The decision fusion in our method is adopted to combine recognition results of each sub-region, which can weaken the influence of poor quality regions resulting from occlusion and illumination variations on the final recognition result of the whole probe image and improve the performance. When the $N$ equals to 81 (9 × 9), the average recognition of four probe sets is the highest. Note that we delete the last row and column of face images in order to uniformly divide them into 81 (9 × 9) sub-images with 11×11 pixels in the AR database in the experiments.

The experimental results of the parameter $N (k \times k)$ on FERET database are shown in Fig. 14. When $l$ equals to 6, it can be seen that the recognition rate goes up promptly from 46% to approximately 82% as $N$ increases. The seventh face image of each person in the FERET database expresses illumination variation. So it further verifies the robustness of our method to illumination. When $l$ is 2, 3, 4 or 5, the curve of the recognition rate rises gently with little range fluctuation. Consequently, the number of $N$ has little effect on pose variation. When $N$ equals to 81 (9 × 9), good performance can be yielded in all cases. Note that we copy the last row and column of face images to add the original images in order to uniformly divide them into 81 (9 × 9) sub-images with 9×9 pixels in the FERET database in the experiments.

We first explore the recognition rates of the internal sub-images under different scales and then further discuss the parameter $M$. Generally, if the size of sub-image is larger, the sub-image contains more information. However, it becomes more sensitive to the variations in expression, illumination, and partial occlusions. Sizes of internal sub-images under Scale $m (0,1,2,\ldots,5)$ on the AR database are 11×11, 15×15, 19×19, 23×23, 27×27, and 31×31 pixels. The experimental results over the AR database are shown in Fig. 15. With the size of sub-image increasing, the recognition rates of four probe sets increase at first, and then keep stable even degrade. Sizes of sub-images under Scale $m (0,1,2,\ldots,4)$ on the FERET database are 9×9, 13×13, 17×17, 21×21, and 25×25 pixels. The recognition results over the FERET database have the similar trend, when the number $l$ of training samples equals to 5 and 6, as shown in Fig. 16. If $l$ equals 2–4, the recognition rate remains rising, as the size increases. So we can conclude that there does not exists a fixed scale to achieve best performance under different variations in expression, pose, illumination, and partial occlusions and features fusion of multi-scale sub-images can improve the robustness.

Experiments on parameter $M$ are performed on AR and FERET databases with $N$ being set to 81 (9 × 9). We extract $M$ sub-images with different sizes based on each of the 49 internal non-overlapped sub-images. $M$ is set to 1–5 respectively on AR database, while $M$ is set to 1–4 on FERET database.

The results of AR database are shown in Fig. 17, where the horizontal axis denotes the number $M$ of sub-images based on the non-overlapped sub-images obtained and the vertical axis describes the recognition rate. The recognition rate of the illumination probe set keeps stable around 93%. For the testing sets of facial expression, occlusion with sunglasses and scarf, the recognition performance increases with $M$ growing. The recognition rate of the scarf probe set grows from 91.83% to 95.33%, when $M$ is 4.

As is shown in Fig. 18, the recognition rates on FERET database go up obviously as $M$ grows in all cases. When $M$ equals to 4, the
recognition rates of \( l = 4, 5 \) are as high as approximate 90%. Generally speaking, the bigger the \( M \) is, the higher the recognition rate is. We can infer that feature extraction based on multi-scale sub-images at the same location contributes to the improvement of face recognition performance with the variations in facial pose, expression and illumination.

### 4.1.3. Performance of the proposed method

In this subsection, we first testify the advantages of the WLD with non-linear quantization rule (NQ) by comparison with the basic WLD with linear quantization rule (LQ). The two neutral images per person in the two sessions of AR database are used as the training set, four different subsets as testing sets. The first \( l \) \((l = 2–6)\) images per person of FERET database are used for training and the remaining for testing. We set \( N \) and \( M \) equal to 81 \((9 \times 9)\) and 4 respectively to obtain the recognition rates of WLD with non-linear quantization rule (NQ) and linear quantization rule (LQ). In order to acquire fair comparisons, we calculate the WLD with linear quantization by setting \( T_1 \) and \( T_2 \) respectively to 8 and 12 to obtain the histogram features with the same length as that of non-linear calculation method. As are shown in Figs. 19 and 20, it is obvious that the WLD calculation method based on non-linear
quantization contributes to the improvement of recognition performance.

In order to further testify the effectiveness of the WLD based on non-linear quantization, it is compared with the LBP. We replace the non-linear quantization WLD with the LBP in our method. The comparison results are shown in Figs. 21 and 22. It is obvious that the WLD with non-linear quantization achieves much better recognition performance than the LBP.

4.2. Computational complexity analysis

In this subsection, we first analyze the computational complexity of WLD and Chi-square similarity measure, and then compute the average time consumption of facial feature extraction with WLD and matching with Chi-square distance respectively on AR and FERET databases. Our experiments using Matlab 7.0 are implemented on Intel Core i7 CPU 2.4 GHz. Give a facial image with size of $X \times W$. The time complexity for WLD can be described as follows [37]:

$$O_{WLD} = C_1XY$$  \hspace{1cm} (13)

where $C_1$ is a constant, which is used for the computation of each pixel in WLD with several additions, divisions and filtering by an arctangent function. The time complexity for Chi-square similarity measure is as follows:

$$O_{Chi} = C_2(N_{sub}(T_1 + T_2))$$  \hspace{1cm} (14)

where $C_2$ is also a constant. We use $C_2$ for the computation of each bin in Chi-square distance through two additions, a multiplication and a division. $N_{sub}$ represents the number of sub-images, which is determined by the parameters $N$ ($k \times k$) and $M$ in our method. $T_1$ and $T_2$ determine the length of the WLD histogram. We set all parameters in our method as the same as Section 4.1. The average time consumptions of feature extraction and matching are shown in Table 3.

4.3. Comparison with other methods

In the experiments on the Yale face database, we set $N$ and $M$ as 81 ($9 \times 9$) and 4 respectively and adopt two evaluation protocols to test our method. Evaluation protocol 1 takes the first $l$ ($l=2$–6) images per person as training and the remaining for testing. The high recognition rates shown in Table 4 demonstrate the effectiveness of our method.

For Evaluation protocol 2, the leave-one-out strategy is applied, that is, each image of one person is removed from the data set to be tested in turns and the others are used for training. The results of the various methods are shown in Table 5, where the results of other methods come from [6]. The Fisherfaces (PCA), ICA and 2DPCA belong to holistic matching methods. The 2DPCA extracts the features based on 2D image matrices has better performance than the PCA and the ICA. Comparing with the ICA, Eigenfaces and 2DPCA, our method achieves the highest recognition rate as 98.18%.

The results of expression, sunglasses and scarf occlusion sets of AR database are listed in Table 6. The recognition results of compared methods are reported in [36]. The subset used in [36] only contains 90 persons (45 men and 45 women) randomly chosen from AR database, while in our experiments the subset including 100 persons (50 men and 50 men) is used to measure our method and the recognition results are shown in Table 6. The local matching methods reported in [36] divide the facial images into overlapped sub-images, and then extract the local feature of each sub-image. The local features based on different locations are combined into a weighted feature vector. The recognition performance of our method is apparently better than the PCA and Fish LDA (FLDA). The combinations of the FLDA and Gabor magnitude information (Gabor-M-FLDA) or phase (Gabor-P-FLDA) information are more robust to variation in expression, but they still achieve poor performance with occlusions. The LBP has better performance than the Gabor feature. The local features combining the LBP and Gabor magnitude (LGBP-M) or phase (LGBP-P) information improve the robustness to variations in occlusions by wearing scarf, but they are still ineffective for the sunglasses. In [36], the magnitude or phase responses by convolving multi-scale and multi-orientation Gabor filters are seen as a third-order Gabor volume and apply LBP on the three orthogonal planes (GV-LBP-TOP-M or GV-LBP-TOP-P) and an more effective calculation method is proposed (E-GV-LBP-M or E-GV-LBP-P), which can achieve good performance except the glasses probe set. Because the sunglasses cover much facial important texture information, most of methods are not able to effectively recognize the facial images with occlusion by wearing sunglasses. The apparent advantage of our method is its effectiveness in sunglass occlusion with the recognition rate more than 40% higher than the peak of other methods. On one hand, the WLD has an excellent ability to describe the image texture information; on the other hand, the feature fusion of multi-scale sub-images at the same locations improves the discriminate power and the decision fusion by

**Table 3** Average time consumptions of feature extraction with WLD and matching with Chi-square distance.

<table>
<thead>
<tr>
<th>Database</th>
<th>Feature extraction (ms)</th>
<th>Matching (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>42.51</td>
<td>2.48</td>
</tr>
<tr>
<td>FERET</td>
<td>37.32</td>
<td>2.18</td>
</tr>
</tbody>
</table>

**Table 4** Recognition results using Evaluation protocol 1 on Yale database.

<table>
<thead>
<tr>
<th>Samples</th>
<th>$l=2$</th>
<th>$l=3$</th>
<th>$l=4$</th>
<th>$l=5$</th>
<th>$l=6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition rate (%)</td>
<td>96.3</td>
<td>97.5</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

**Table 5** Performance comparison with other methods using Evaluation protocol 2 on Yale database.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Leave-one-out (%)</td>
<td>71.52</td>
<td>71.52</td>
<td>84.24</td>
<td>98.18</td>
</tr>
</tbody>
</table>

**Table 6** Performance comparison with other methods on AR database.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Expression (%)</th>
<th>Sunglass (%)</th>
<th>Scarf (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA [36]</td>
<td>74.07</td>
<td>12.96</td>
<td>2.41</td>
</tr>
<tr>
<td>FLDA [36]</td>
<td>72.41</td>
<td>11.85</td>
<td>9.81</td>
</tr>
<tr>
<td>Gabor-M-FLDA [36]</td>
<td>86.30</td>
<td>21.11</td>
<td>31.48</td>
</tr>
<tr>
<td>Gabor-P-FLDA [36]</td>
<td>77.04</td>
<td>28.89</td>
<td>55.37</td>
</tr>
<tr>
<td>LBP [36]</td>
<td>87.24</td>
<td>34.63</td>
<td>47.04</td>
</tr>
<tr>
<td>LGBP-M [36]</td>
<td>86.11</td>
<td>37.59</td>
<td>82.59</td>
</tr>
<tr>
<td>LGBP-P [36]</td>
<td>85.93</td>
<td>37.04</td>
<td>83.52</td>
</tr>
<tr>
<td>E-GV-LBP-M [36]</td>
<td>90.93</td>
<td>42.72</td>
<td>82.78</td>
</tr>
<tr>
<td>E-GV-LBP-P [36]</td>
<td>89.81</td>
<td>44.07</td>
<td>86.67</td>
</tr>
<tr>
<td>GV-LBP-TOP-M [36]</td>
<td>90.56</td>
<td>53.89</td>
<td>87.41</td>
</tr>
<tr>
<td>GV-LBP-TOP-P [36]</td>
<td>91.11</td>
<td>46.11</td>
<td>90.37</td>
</tr>
<tr>
<td>Ours</td>
<td>96</td>
<td>95.33</td>
<td>96.67</td>
</tr>
</tbody>
</table>
matching results based on different locations voting reduces the impact of sunglasses.

Performance comparisons are made between the proposed method (with the recognition rates when \(N=81\) and \(M=4\)) and some other methods referring to [42], such as the PCA, LDA, LPP, Laplacian PCA (LPCA), linear Laplacian discrimination (LLD), and multi-manifold discriminant analysis (MMDA), as shown in Table 7. The LPCA is an extension of the PCA by optimizing the locally weighted scatters instead of the single global non-weighted scatter in the PCA [43]. The LLD is an extension of the LDA, which defines the within-class scatter and the between-class scatter using similarities based on pairwise distances in sample spaces [44]. The MMDA defines two different graphs, the within-class graph and the between-class graph, to characterize the within-class compactness and the between-class separability and finds a matrix simultaneously maximizing the between-class and minimizing the within-class scatter [42]. The LPCA, the LLD, and the MMDA are the holistic matching methods and achieve much better performance than the PCA, the LDA and the LPP. It is apparently indicated that our method has the best performance with high recognition rate in general ranging from 80.10\% to 89.83\%.

As are shown in Table 6, our method is of good robustness to partial occlusions by wearing scarf and sunglasses. The experiments are conducted to further testify its robustness to variations in contiguous occlusions on AR database. In our experiments, we replace a randomly located square block of each image in expression subset with an unrelated image as in [45] to construct the test set. The square block is set to 20×20, 40×40, and 60×60 respectively as shown in Fig. 23. The two neutral images per person still sever as training samples. Then we test the performance of sparse representation based classification (SRC) and partitioned SRC algorithm [45] with the same training and test sets. The SRC represents the test sample as a sparse linear combination of training samples and makes use of \(l^1\) minimization technique to compute the sparse coefficients. The partitioned SRC divides the facial image into blocks, processes each block independently applying the SRC, and aggregates their results by voting. In the experiments, the partitioned SRC divides facial images into sub-images with the size of 20×30 as reported in [45]. The comparison results are shown in Table 8. The SRC and partitioned SCR can achieve high recognition rates when the training samples are enough, while their performances degrade obviously in our experiments with only two images per person for training. The results show that our method is still of good robustness to variations in random contiguous occlusions.

5. Conclusion

This paper proposes a method for face recognition using multi-scale WLD and multi-level information fusion approaches. Experimental results show that our method can achieve good recognition performance. In summary, there are three contributions: (1) The WLD, a powerful and robust local descriptor, is introduced to represent the facial images. It has an excellent ability to capture local intensity variations in facial images by its two components. Moreover, we modify it by a non-linear quantization rule to enhance its discriminative power. (2) Multi-scale sub-images at the same location are employed to extract more discriminative features. It is beneficial to describe different facial local structures and improves the robustness to variations in facial expressions and poses. It also can be seen to increase the number of the training samples. (3) Decision fusion of matching results based on different locations can reduce the negative influence of poor quality of local regions in face images coming from the variations in illumination and partial occlusion.

Table 7

<table>
<thead>
<tr>
<th>Methods</th>
<th>(l=2) (%)</th>
<th>(l=3) (%)</th>
<th>(l=4) (%)</th>
<th>(l=5) (%)</th>
<th>(l=6) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA [42]</td>
<td>47.60</td>
<td>41.50</td>
<td>50.17</td>
<td>40.00</td>
<td>30.50</td>
</tr>
<tr>
<td>LDA [42]</td>
<td>66.00</td>
<td>59.25</td>
<td>70.67</td>
<td>70.00</td>
<td>59.00</td>
</tr>
<tr>
<td>LPP [42]</td>
<td>40.40</td>
<td>35.75</td>
<td>46.67</td>
<td>61.25</td>
<td>67.00</td>
</tr>
<tr>
<td>LPCA [42]</td>
<td>52.70</td>
<td>49.25</td>
<td>61.33</td>
<td>68.75</td>
<td>72.00</td>
</tr>
<tr>
<td>LLD [42]</td>
<td>63.50</td>
<td>57.00</td>
<td>70.33</td>
<td>71.25</td>
<td>71.50</td>
</tr>
<tr>
<td>MMDA [42]</td>
<td>67.20</td>
<td>59.75</td>
<td>72.33</td>
<td>72.50</td>
<td>75.00</td>
</tr>
<tr>
<td>Ours</td>
<td>80.10</td>
<td>75.50</td>
<td>89.83</td>
<td>89.75</td>
<td>88.00</td>
</tr>
</tbody>
</table>

Table 8

<table>
<thead>
<tr>
<th>Methods</th>
<th>0 (%)</th>
<th>20×20 (%)</th>
<th>40×40 (%)</th>
<th>60×60 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>78.00</td>
<td>76.50</td>
<td>69.67</td>
<td>49.50</td>
</tr>
<tr>
<td>Partitioned SRC</td>
<td>65.80</td>
<td>61.33</td>
<td>51.83</td>
<td>36.67</td>
</tr>
<tr>
<td>Ours</td>
<td>96.00</td>
<td>95.83</td>
<td>91.17</td>
<td>82.83</td>
</tr>
</tbody>
</table>

Fig. 23. Face images of one person with random contiguous occlusions on AR database.
Acknowledgments

This work is the National Basic Research Program of China (Youth 973 Program) (Grant No. 2013CB336500) supported by the Natural National Science Foundation of China (No. 6172161 and No. 61172143).

References


Shutao Li received the B.S., M.S., and Ph.D. degrees in computer engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively. In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, U.K., working with Prof. J.S. Taylor. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 160 refereed papers. His professional interests are information fusion, pattern recognition, bioinformatics, and image processing. Dayi Gong was born in Hunan, China, in 1987. He received his B.S. degree in automation from Hunan University, China, in 2010. He is now pursuing his M.S. degree in control science and engineering with the College of Electrical and Information Engineering, Hunan University. His research interests include image processing and pattern recognition.

Dayi Gong was born in Hunan, China, in 1987. He received his B.S. degree in automation from Hunan University, China, in 2010. He is now pursuing his M.S. degree in control science and engineering with the College of Electrical and Information Engineering, Hunan University. His research interests include image processing and pattern recognition.

Yuan Yuan is a full professor with the Chinese Academy of Sciences (CAS), China. He has made major research contributions in the areas of various multimedia data processing and machine learning. His major research interests include Visual Information Processing and Image/Video Content Analysis. She has published over a hundred papers, including over 60 in reputable journals, like IEEE Transactions and Pattern Recognition, as well as conference papers in CVPR, BMVC, ICFP, ICASSP, etc.