Face Recognition Based on Pixel-Level and Feature-Level Fusion of the Top-Level's Wavelet Sub-Bands

Zheng-Hai Huang[a,b], Wen-Juan Li[a], Jun Wang[a,c], Ting Zhang[a,*]

[a] *Center for Applied Mathematics of Tianjin University, Tianjin 300072, P.R. China*
[b] *Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, P.R. China*
[c] *Center for Combinatorics, Nankai University, Tianjin 300071, P.R. China*

Abstract

The traditional wavelet-based approaches directly use the low frequency sub-band of wavelet transform to extract facial features. However, the high frequency sub-bands also contain some important information corresponding to the edge and contour of face, reflecting the details of face, especially the top-level's high frequency sub-bands. In this paper, we propose a novel technique which is a joint of pixel-level and feature-level fusion at the top-level's wavelet sub-bands for face recognition. We convert the problem of finding the best pixel-level fusion coefficients of high frequency wavelet sub-bands to two optimization problems with the help of principal component analysis and linear discriminant analysis, respectively; and propose two alternating direction methods to solve the corresponding optimization problems for finding transformation matrices of dimension reduction and optimal fusion coefficients of wavelet high frequency sub-bands. The proposed methods make full use of four top-level's wavelet sub-bands rather than low frequency sub-band only. Experiments are carried out on the FERET, ORL and AR face databases, which indicate that our methods are effective and robust.

Keywords: Face recognition, wavelet transform, pixel-level fusion, feature -level fusion, principle component analysis, linear discriminant analysis

[*]Corresponding author. Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, P.R. China. Email address: zt6906982@tju.edu.cn (T. Zhang). Tel.: 86-22-27403615, fax: 86-22-27403615.

1. Introduction

Face recognition, as a biological feature recognition, has been one of the most active research areas in computer vision, pattern recognition and biometrics [1]. Face recognition has several advantages over other biometric modalities such as fingerprint and iris. Besides being natural and nonintrusive, the most important advantage of face recognition is that it can be captured at a distance and in a friendly manner. Various face recognition algorithms have been devised in the literature. However, up to now, face recognition is still faced with a number of challenges such as varying illumination, facial expression and poses [2, 3, 4].

Feature extraction and classification are two key steps of a face recognition system. Feature extraction provides an effective representation of face images to decrease the computational complexity of the classifier, which can greatly enhance the performance of a face recognition system; while classification is to distinguish those features with a good classifier. Therefore, in order to improve the recognition rate of a face recognition system, it is crucial to find a good feature extractor and an effective classifier [5]. In this paper, we focus on feature extraction methods by using wavelet transform with the help of the classical principal component analysis (PCA) [6] and linear discriminant analysis (LDA) [7].

Wavelet transform, which is an increasingly popular tool in image processing and computer vision, has been investigated in many applications, such as compression, detection, recognition, image retrieval et al., due to its great advantages with the nice features of space-frequency localization and multiresolution. Researchers have developed face recognition algorithms by combining discrete wavelet transform (DWT) with other methods (see, for example, [8, 9, 10, 11, 12, 13] and references therein). Through a two-dimensional DWT (2D-DWT), an image of face is transformed into two parts: a low frequency sub-band and three high frequency sub-bands. On one hand, the low frequency sub-band plays a dominant role in four sub-bands for approximation of the original image; and on the other hand, when the expression changes or the image is affected by occlusion, the low frequency sub-band has no obvious change but the high frequency sub-bands change obviously, so the low frequency sub-band is usually single-handed used for identification in image recognition [10]. The authors in [14] directly used the low frequency sub-band of wavelet transform to extract facial features. However, though the high frequency sub-bands don't contain as much information as the low frequency sub-band, they also contain important information corresponding to the edge and contour of face, reflecting the details of face [15]. Moreover, we implement the classical DWT+PCA on the ORL database by using 4-level wavelet transform and every top-level's high frequency sub-band is single-handed used for identification in face recognition. The obtained recognition rates by three top-level's high frequency sub-bands are 67.5%, 61.0% and 43.0%, respectively. This implies that every top-level's high frequency sub-band contains much information of face features. Thus, it is more reasonable to use the information of more sub-bands for face recognition. A natural question is *how to use the information of these sub-bands effectively for face recognition*?

In recent years, data fusion has been developed rapidly and widely applied in many areas such as object recognition, pattern classification, image processing, and so on. Generally speaking, data fusion is performed at three different processing levels according to the stage at which the fusion takes place: pixel-level, feature-level and decision-level [16]. In pixel-level fusion, the information derived from multiple feature sets is assimilated and integrated into a final decision directly. Many fusion algorithms for the pixel-level fusion have been proposed, from the simplest weighted averaging to more advanced multiscale methods. There are two existing feature-level fusion methods. One is to group two sets of feature vectors into a union-vector, and another one is to combine two sets of feature vectors by a complex vector. Both feature fusion methods can increase the recognition rate. The advantage of feature-level fusion lies in two aspects: firstly, it can derive the most discriminatory information from original multiple feature sets; secondly, it is able to eliminate the redundant information resulting from the correlation between distinct feature sets. The decision-level fusion, delegated by multi-classifier combination, has been one of the hot research fields on pattern recognition, and has achieved successful application in face recognition.

In this paper, based on 2D-DWT, we propose a joint of pixel-level and feature-level fusion at the top-level's wavelet sub-bands technique which can make full use of four top-level's wavelet sub-bands, and we abbreviate it as TWSBF. The proposed technique is different from those image fusion techniques developed in the literature (see, for example, [17, 18, 19, 20] and references therein). Firstly, we conduct dimension reduction on low frequency wavelet sub-band. Then, we consider the top-level's three high frequency sub-bands, in which we apply the pixel-level fusion. In this way, we can keep more discriminatory features and avoid redundancy. We convert the problem of finding the best fusion coefficients to two optimization problems based on PCA and LDA, respectively; and propose two alternating direction methods to solve the corresponding optimization problems for finding the optimal transformation matrices of dimension reduction and optimal fusion coefficients of wavelet sub-bands. Finally, we process feature-level fusion on low frequency sub-band after dimension reduction and fused high frequency vector. The experiments are carried out on the FERET, ORL and AR face databases. The numerical experimental results demonstrate that the proposed methods possess higher recognition rates than the classical wavelet-based methods and some popular methods at present.

The rest of this paper is organized as follows. In Section 2, after a brief introduction of 2D-DWT, we propose the model of TWSBF. The model and algorithm based on TWSBF and PCA are investigated in Section 3; and the model and algorithm based on TWSBF and LDA are discussed in Section 4. The numerical experimental results on the FERET, ORL and AR face image databases are reported in Section 5. The final remarks are given in Section 6.

2. Model of TWSBF

Let $L^2(R)$ denote the square integrable space. The continuous wavelet transform of a one-dimensional function $f(t) \in L^2(R)$ is defined as

$$W_f(a,b) = \int_R f(t)\overline{\psi_{a,b}(t)}dt, \tag{1}$$

where the wavelet basis functions $\psi_{a,b}(t)$ can be expressed as

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}}\psi(\frac{t-b}{a}),$$

in which $\psi(t)$ is called mother wavelet and the parameters a and b stand for the scale and position, respectively. Equation (1) can be discretized by imposing restrictions on a and b with $a = 2^n$ and $b \in Z$.

The DWT of a one-dimensional signal is processed by transforming it into two parts with a low-pass filter and a high-pass filter. The low frequency part is split again into two parts of high and low frequencies [21]. In image processing, similar to the one-dimensional DWT, the DWT for a two-dimensional image can be constructed by applying the one-dimensional

DWT at horizontal and vertical directions, respectively. The process of 2D-DWT is shown in Figure 1, in which an image is first filtered in the horizontal direction with low-pass and high-pass filters. Then the filtered outputs are downsampled by a factor of 2. Moreover, the same process is applied in the

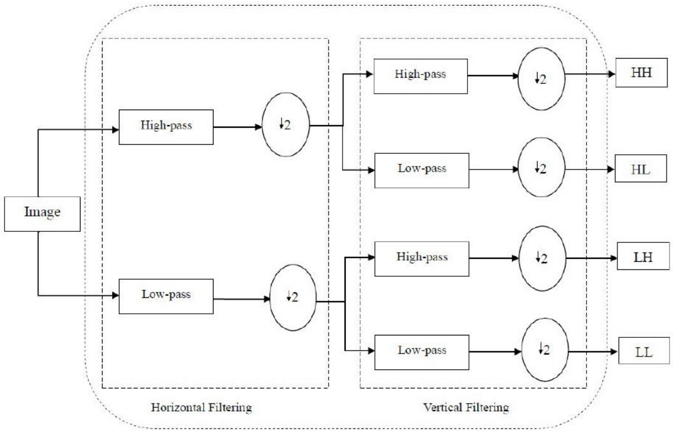
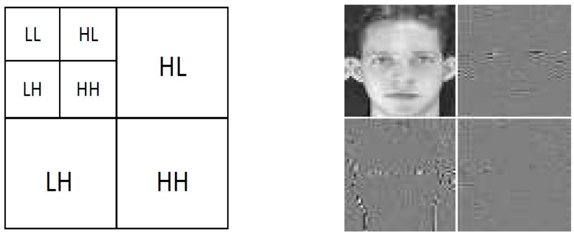


Figure 1: The Flowchart of 1-Level 2D-DWT

vertical direction. Thus, an image is decomposed into 4 sub-bands denoted by LL, HL, LH, HH. Each sub-band can be thought of a smaller version of the image representing different properties. The sub-band LL is a coarser approximation to the original image; the sub-bands LH and HL record the changes of the image along horizontal and vertical directions, respectively; and the sub-band HH shows the changes of the image along diagonal direction. The sub-bands LH, HL and HH are all the high frequency components of the image. Further decomposition can be conducted on the LL sub-band. The 1-level 2D-DWT means that an original image is decomposed into a low frequency sub-band and three high frequency sub-bands. Generally, the t-level 2D-DWT ($t > 1$) means that the low frequency sub-band obtained in the $(t-1)$-level 2D-DWT is further decomposed into a low frequency sub-band and three high frequency sub-bands. Figure 2(a) shows the flowchart of a 2-level 2D-DWT and Figure 2(b) shows the face image with a 1-level 2D-DWT.

Generally, sub-band LL is low frequency component of the image which contains most information of the original image; and sub-bands HL, LH and HH stand for the high-frequency components of the image which reflect the details of images. Since the low frequency sub-band plays a dominant



(a) The 2-Level 2D-DWT Diagram

(b) The 1-Level 2D-DWT of a Face Image

Figure 2: The Demonstration of 2D-DWT

role in four sub-bands for approximation of the original image, if a t-level wavelet-based method is used, then the low frequency sub-band obtained by the t-th-level wavelet transform (i.e., the top-level's low frequency sub-band) is usually single-handed used for identification in image recognition in the literature. However, though the high frequency sub-bands don't contain as much information as the low frequency sub-band, they also contain much important information. Thus, it is more reasonable to use the information of all top-level's sub-bands for face recognition. Based on these analysis, we now propose a new wavelet-based technique (TWSBF) for feature extraction, which makes full use of all top-level's wavelet sub-bands. Such a technique is described as follows.

Let $I(x, y)$ be a face image, by using 1-level or multi-level wavelet transform to $I(x, y)$, we get the top-level's low frequency sub-band matrix $LL \in R^{m\times n}$ and high frequency sub-bands matrices $HL, LH, HH \in R^{m\times n}$. Without loss of generality, let $L, H, V, D \in R^l$ represent the vectorization of LL, HL, LH, HH, where $l = m \times n$. Supose the dimension reduction function of L is f_L, where the function f_L can be PCA, LDA and so on. Then the low dimensional representation of L is

$$DL = f_L(L) \in R^{d_1},$$

where d_1 is the number of dimension reduction, which can be seen the best representation of L.

While the hight frequency sub-bands of $I(x, y)$ can be further expressed as an $l \times 3$ matrix: $A = [H\ V\ D] \in R^{l\times 3}$. Therefore, the linear combination of the high frequency sub-bands can be described as follows:

$$S = u_1H + u_2V + u_3D = Au \in R^l \tag{2}$$

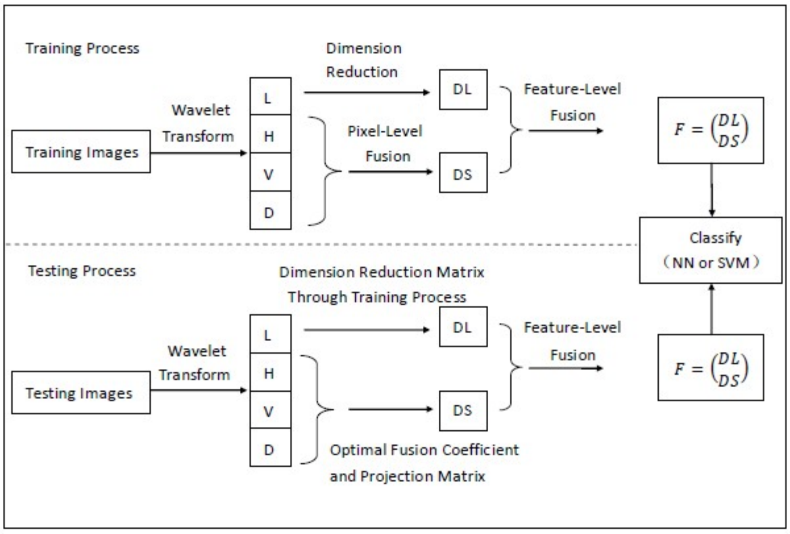


Figure 3: The Flow Chart of TWSBF

where $u = (u_1, u_2, u_3)^T \in R^3$ is a coefficient vector of the high frequency wavelet sub-band fusion, which gives a model of pixel-level fusion of three top-level's high frequency wavelet sub-bands.

In order to make such a fusion technique of wavelet sub-bands more effective, it is necessary to find the optimal coefficient vector u in (2) so that S is the best representation of A for face recognition. Suppose that f_H is the function of this process, and S is expressed as

$$DS = f_H(S) = f_H(Au) \in R^{d_2},$$

where d_2 is the number of fused dimension.

By far, we get the low frequency and the high frequency representations of I, i.e., DL and DS. For the application of high frequency and low frequency information' difference, feature-level fusion was used. That is we use

$$F = \begin{pmatrix} DL \\ DS \end{pmatrix} = \begin{pmatrix} f_L(L) \\ f_H(S) \end{pmatrix} \in R^{d_1+d_2}$$

to represent the face $I(x, y)$. Figure 3 vividly illustrates this fusion flow. While the appropriate function f_H is focused in this article.

3. TWSBF plus PCA

3.1. Classical PCA

The feature extraction is to transform data from a high dimensional space to a lower dimensional space, which can be linear or nonlinear. PCA [6] is a linear dimension reduction technique which extracts the desired principal components of the multidimensional data.

Suppose that the set $\{x_1, x_2, \cdots, x_m\}$ with $x_i \in R^l$ comes from m training samples, then PCA is trying to find a transformation matrix $W \in R^{l\times\gamma}$ such that $y_i = W^T(x_i - \overline{x})$, where $y_i \in R^\gamma$ and $\overline{x} = \frac{1}{m}\sum_{i=1}^m x_i$ is the mean of the samples. To make y_i be on behalf of x_i in the projection subspace, the covariance of y_i must be as large as possible. The covariance matrix of y_i in projection subspace is defined by

$$\begin{aligned} S &= \sum_{i=1}^{m} y_i y_i^T \\ &= \sum_{i=1}^{m} \left[W^T(x_i - \overline{x})\right]\left[W^T(x_i - \overline{x})\right]^T \\ &= W^T\left[\sum_{i=1}^{m}(x_i - \overline{x})(x_i - \overline{x})^T\right]W \\ &= W^T S_{PCA} W, \end{aligned}$$

where $S_{PCA} = \sum_{i=1}^m (x_i - \overline{x})(x_i - \overline{x})^T$ is the covariance matrix of training samples. So, the classical PCA needs to solve the following optimization problem to obtain the transformation matrix W:

$$\max_{W}\ tr(W^T S_{PCA} W), \tag{3}$$

where $tr(\cdot)$ is the trace of a matrix. From the property of Rayleigh quotient [22], the solution of (3) can be obtained by solving the following eigenvalue problem:

$$S_{PCA} w = \lambda w. \tag{4}$$

Suppose that $w_1, w_2, \cdots, w_\gamma$ are eigenvectors of (4) which correspond to the γ largest eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_\gamma$ satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\gamma$, then the solution of (3) is $W^* = [w_1\ w_2\ \cdots\ w_\gamma]$.

3.2. Model of TWSBF plus PCA

In this subsection, we combine TWSBF with PCA to propose a new model for feature extraction. Let m be the number of training face images. For any $i = 1, \cdots, m$, the top-level's wavelet sub-bands of the i-th image are L_i, H_i, V_i, D_i, respectively. Firstly, we implement dimension reduction on low frequency sub-band L_i through PCA and obtain DL_i. Then we implement pixel-level fusion on high frequency sub-bands as follows.

Denote $A_i = [H_i\ V_i\ D_i]$. The mean high frequency wavelet image of all training images is $\overline{A} = \frac{1}{m}\sum_{i=1}^{m} A_i$. The pixel-level fusion of the high frequency wavelet image A_i is given by

$$S_i = u_1 H_i + u_2 V_i + u_3 D_i = A_i u \in R^l,$$

where $u = (u_1, u_2, u_3)^T \in R^3$ is a coefficient vector of wavelet high frequency sub-bands fusion.

Let $\overline{S}$ be the total mean vector, i.e., $\overline{S} = \overline{A}u$; and y_i be the projection vector of S_i, i.e., $y_i = W^T(S_i - \overline{S})$. Then the covariance matrix of projection vector is defined by

$$\begin{aligned} S(W,u) &= \sum_{i=1}^{m} y_i y_i^T \\ &= \sum_{i=1}^{m} \left[W^T(S_i - \overline{S})\right]\left[W^T(S_i - \overline{S})\right]^T \\ &= W^T \left[\sum_{i=1}^{m} (S_i - \overline{S})(S_i - \overline{S})^T\right] W \\ &= W^T \left[\sum_{i=1}^{m} (A_i u - \overline{A}u)(A_i u - \overline{A}u)^T\right] W \\ &= W^T \left[\sum_{i=1}^{m} (A_i - \overline{A})uu^T(A_i - \overline{A})^T\right] W. \end{aligned} \tag{5}$$

If we define the covariance matrix of high-frequency wavelet sub-bands by

$$S_c(u) = \sum_{i=1}^{m} (A_i - \overline{A})uu^T(A_i - \overline{A})^T \in R^{l\times l}, \tag{6}$$

then (5) becomes

$$S(W,u) = W^T S_c(u) W.$$

Therefore, we need to solve the following optimization problem to obtain the optimal transformation matrix W and the optimal coefficient vector u:

$$\max_{W,u} tr(W^T S_c(u)W). \tag{7}$$

Comparing the classical PCA model (3) with the new model (7), it is easy to see that the new model (7) is more difficult to solve than the classical PCA model (3), since, in the new model (7), we need to find the optimal coefficient vector u besides the transformation matrix W. Fortunately, the new model (7) is computationally tractable, which is investigated in the following subsection.

3.3. Algorithm for Model of TWSBF plus PCA

In order to design the effective method to solve the optimization problem (7), the following property is a key.

Theorem 3.1. *Denote*

$$L_c(W) = \sum_{i=1}^{m}(A_i - \overline{A})^T WW^T(A_i - \overline{A}) \in R^{3\times 3}. \tag{8}$$

Then, we have

$$tr(W^T S_c(u)W) = tr(u^T L_c(W)u).$$

Proof: By using the property of the exchange of trace, we can obtain

$$\begin{aligned} tr(W^T S_c(u)W) &= tr\left(\sum_{i=1}^{m} W^T(A_i - \overline{A})uu^T(A_i - \overline{A})^T W\right) \\ &= tr\left(\sum_{i=1}^{m} u^T(A_i - \overline{A})^T WW^T(A_i - \overline{A})u\right) \\ &= tr\left(u^T\left(\sum_{i=1}^{m}(A_i - \overline{A})^T WW^T(A_i - \overline{A})\right)u\right) \\ &= tr(u^T L_c(W)u), \end{aligned}$$

which completes the proof.

It is easy to see that

- if u is fixed, the optimal transformation matrix W in (7) can be obtained by the similar method for solving the classical PCA model; and
- if W is fixed, we know from Theorem 3.1 and Rayleigh quotient [22] that the optimal coefficient vector u in (7) is the corresponding eigenvector of the largest eigenvalue of matrix $L_c(W)$.

Thus, we can design an alternating direction method to solve the new model (7), which is described in Algorithm 3.1.

Algorithm 3.1. *(**TWSBF+PCA**)*

1: *Let L_i, H_i, V_i, D_i be four sub-bands obtained by the t-level 2D-DWT.*

2: *Implement PCA dimension reduction on L_i and obtain DL_i.*

3: *Denote $A_i = [H_i\ V_i\ D_i]$. Set $k = 0$, and provide an initial value for u, i.e., $u = u^{[k]}$.*

4: *Compute $S_c(u)$ by (6) , and its eigenvector w_i corresponding to the i-th largest eigenvalue of $S_c(u)$, where $i = 1, 2, \cdots, \gamma$. Let $W^{[k+1]} = [w_1\ w_2\ \cdots\ w_\gamma]$.*

5: *Compute $L_c(W^{[k+1]})$ by (8), and its eigenvector $u^{[k+1]}$ corresponding to the largest eigenvalue of $L_c(W^{[k+1]})$.*

6: *If $k > 0$ and*

$$\left|tr(S(W^{[k+1]}, u^{[k+1]})) - tr(S(W^{[k]}, u^{[k]}))\right| < \varepsilon,$$

then

Let $u^* = u^{[k+1]}$, $W^* = W^{[k+1]}$. Stop.

else

Update $u = u^{[k+1]}$, $W = W^{[k+1]}$, $k = k + 1$. Go to 3.

end if

7: *Compute DS_i through u^* and W^*. Then implement feature-level fusion on DL_i and DS_i and obtain F_i.*

4. TWSBF plus LDA

4.1. Classical LDA

From Section 3, we know that PCA is a dimension reduction method without discrimination because there is no use of the sample label information during the discussion of PCA. However, the dimension reduction method with discrimination is more advantageous for classification problems in general. LDA is a widely used criterion of dimension reduction method whose main idea is to make the samples' distribution in the same class in the low dimensional space compact as much as possible (minimize *within-class* scatter); and make the distance of samples in different classes in the low dimensional space separate as much as possible (maximize *between-class* scatter) [23, 24].

Given c training classes of faces. Suppose that each class i has m_i training face images and m is the total number of face images satisfying $\sum_{i=1}^{c} m_i = m$. Let x_{ij} be the vector from j-th training image in class i. LDA is trying to find a transformation matrix $W \in R^{l\times\gamma}$ such that $y_{ij} = W^T x_{ij}$ where $y_{ij} \in R^{\gamma}$, which maximizes the *between-class* scatter matrix of the projected training samples and minimizes the *within-class* scatter matrix of the projected training samples. Suppose that $\overline{x}_i$ denotes the mean vector of class i, i.e., $\overline{x}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij}$; and $\overline{x}$ denotes the mean vector of the whole training set, i.e., $\overline{x} = \frac{1}{m}\sum_{i=1}^{c}\sum_{j=1}^{m_i} x_{ij}$. Then, the *between-class* and *within-class* scatter matrices in the original training space are given, respectively, by

$$\begin{aligned} S_b &= \sum_{i=1}^{c} \frac{m_i}{m}(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})^T, \\ S_w &= \sum_{i=1}^{c} \frac{m_i}{m}\left[\frac{1}{m_i - 1}\sum_{j=1}^{m_i}(x_{ij} - \overline{x}_i)(x_{ij} - \overline{x}_i)^T\right], \end{aligned}$$

and hence, the *between-class* and *within-class* scatter matrices in the projection subspace are given, respectively, by

$$\begin{aligned} \widetilde{S}_b &= \sum_{i=1}^{c} \frac{m_i}{m}(\overline{y}_i - \overline{y})(\overline{y}_i - \overline{y})^T \\ &= \sum_{i=1}^{c} \frac{m_i}{m}(W^T\overline{x}_i - W^T\overline{x})(W^T\overline{x}_i - W^T\overline{x})^T \\ &= W^T\left[\sum_{i=1}^{c} \frac{m_i}{m}(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})^T\right] W \end{aligned}$$

$$
\begin{aligned}
&= W^T S_b W, \\
\widetilde{S}_w &= \sum_{i=1}^{c} \frac{m_i}{m} \left[\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - \overline{y}_i)(y_{ij} - \overline{y}_i)^T \right] \\
&= \sum_{i=1}^{c} \frac{m_i}{m} \left[\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (W^T x_{ij} - W^T \overline{x}_i)(W^T x_{ij} - W^T \overline{x}_i)^T \right] \\
&= W^T \left\{ \sum_{i=1}^{c} \frac{m_i}{m} \left[\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (x_{ij} - \overline{x}_i)(x_{ij} - \overline{x}_i)^T \right] \right\} W \\
&= W^T S_w W.
\end{aligned}
$$

So, the classical LDA needs to solve the following optimization problem:

$$
\max_{W} \frac{tr(W^T S_b W)}{tr(W^T S_w W)}. \tag{9}
$$

From the property of Rayleigh quotient [22], the solution of (9) can be obtained by solving the following generalized eigenvalue problem:

$$
S_b w_i = \lambda S_w w_i.
$$

Suppose that $\{w_i | i = 1, 2, \cdots, \gamma\}$ is the set of generalized eigenvectors of matrices S_b and S_w corresponding to the γ largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \cdots, \gamma\}$. Then the solution of (9) is $W^* = [w_1 \ \cdots \ w_\gamma]$.

4.2. Model of TWSBF plus LDA

Given a set of training face images with class labels, we can generate a pixel-level combined feature-level fusion of the wavelet sub-bands for each image. In order to achieve the better recognition performance, we combine TWSBF with the classical LDA to propose a new model for feature extraction.

Let c be the number of training face classes, and $L_{ij}, H_{ij}, V_{ij}, D_{ij}$ be the top-level's wavelet sub-bands of j-th face image in class i, where $i = 1, 2, \cdots, c$, $j = 1, 2, \cdots, m_i$, and m_i denotes the number of training samples in class i. Firstly, we implement dimension reduction in low frequency sub-band L_{ij} through LDA and obtain DL_{ij}. Then we implement pixel-level fusion on high frequency sub-bands as follows.

Let $A_{ij} = [H_{ij} \ V_{ij} \ D_{ij}]$ be the top-level's high frequency wavelet sub-band matrix of j-th face image in class i, The mean high frequency wavelet image of training images in class i is $\overline{A_i} = \frac{1}{m_i}\sum_{j=1}^{m_i} A_{ij} = [\overline{H_i}\ \overline{V_i}\ \overline{D_i}]$ and the mean high frequency wavelet image of all training images is $\overline{A} = \frac{1}{m}\sum_{i=1}^{c}\sum_{j=1}^{m_i} A_{ij} = [\overline{H}\ \overline{V}\ \overline{D}]$, where m is the total number of training images satisfying $m = \sum_{i=1}^{c} m_i$. Suppose that the pixel-level fusion of three high frequency wavelet sub-bands of $A_{ij} = [H_{ij}\ V_{ij}\ D_{ij}]$ is S_{ij}, i.e.,

$$S_{ij} = u_1 H_{ij} + u_2 V_{ij} + u_3 D_{ij} = [H_{ij}\ V_{ij}\ D_{ij}]u = A_{ij}u \in R^l,$$

where $u = (u_1, u_2, u_3)^T \in R^3$ is a coefficient vector of high frequency wavelet sub-band fusion. Denote the mean vector in class i by $\overline{S_i} = \overline{A_i}u$ and the total mean vector by $\overline{S} = \overline{A}u$. Then, the *between-class* scatter matrix $S_b(u)$ and the *within-class* scatter matrix $S_w(u)$ are defined, respectively, by

$$\begin{aligned} S_b(u) &= \sum_{i=1}^{c} P_i(\overline{S_i} - \overline{S})(\overline{S_i} - \overline{S})^T \\ &= \sum_{i=1}^{c} P_i[(\overline{A_i} - \overline{A})u][(\overline{A_i} - \overline{A})u]^T \\ &= \sum_{i=1}^{c} P_i[(\overline{A_i} - \overline{A})uu^T(\overline{A_i} - \overline{A})^T], \qquad (10) \\ S_w(u) &= \sum_{i=1}^{c} P_i\left[\frac{1}{m_i - 1}\sum_{j=1}^{m_i}(S_{ij} - \overline{S_i})(S_{ij} - \overline{S_i})^T\right] \\ &= \sum_{i=1}^{c} P_i \frac{1}{m_i - 1}\sum_{j=1}^{m_i}[(A_{ij} - \overline{A_i})u][(A_{ij} - \overline{A_i})u]^T \\ &= \sum_{i=1}^{c} P_i \frac{1}{m_i - 1}\sum_{j=1}^{m_i}[(A_{ij} - \overline{A_i})uu^T(A_{ij} - \overline{A_i})^T], \qquad (11) \end{aligned}$$

where P_i is the prior probability of class i which is generally calculated with $P_i = m_i/m$. From the analysis of Subsection 4.1, we need to maximize the following function:

$$J(W, u) = \frac{tr(W^T S_b(u) W)}{tr(W^T S_w(u) W)}, \qquad (12)$$

where W is an $l \times \gamma$ transformation matrix, which is formed by a set of projection basis vectors $w_1, w_2, \cdots, w_\gamma$.

By the property of Rayleigh quotient [22], maximizing the function given in (12) is equivalent to solving the following optimization problem:

$$\max_{W,u} \operatorname{tr}(W^T S_b(u) W) \quad \text{s.t.} \quad W^T S_w(u) W = I, \tag{13}$$

where I denotes the identity matrix of $\gamma \times \gamma$.

4.3. Algorithm for Model of TWSBF plus LDA

In order to solve the optimization problem (13), we need the following key property, whose proof is similar to the one of Theorem 3.1. Thus, we omit the proof here.

Theorem 4.1. *Denote*

$$\begin{aligned} L_b(W) &= \sum_{i=1}^{c} P_i[(\overline{A_i} - \overline{A})^T W W^T (\overline{A_i} - \overline{A})] \in R^{4\times 4}, & (14)\\ L_w(W) &= \sum_{i=1}^{c} P_i \frac{1}{m_i - 1} \sum_{j=1}^{m_i} [(A_{ij} - \overline{A_i})^T W W^T (A_{ij} - \overline{A_i})] \in R^{4\times 4}. & (15) \end{aligned}$$

Then, we have

$$\operatorname{tr}(W^T S_b(u) W) = u^T L_b(W) u \quad \text{and} \quad \operatorname{tr}(W^T S_w(u) W) = u^T L_w(W) u.$$

Thus, the problem (13) can be solved by the following approach.

Case 1. Suppose that u is fixed. In this case, the optimal transformation matrix W in (13) can be obtained by the similar method for solving the classical LDA. However, the *within-class* scatter matrix $S_w(u)$ is usually singular or close to singular. To improve the performance of the proposed method, we use the following strategy of PCA plus LDA [7]. Define the total scatter matrix $S_T(u)$ of PCA by

$$\begin{aligned} S_T(u) &= \frac{1}{m} \sum_{i=1}^{c} \sum_{j=1}^{m_i} (S_{ij} - \overline{S})(S_{ij} - \overline{S})^T \\ &= \frac{1}{m} \sum_{i=1}^{c} \sum_{j=1}^{m_i} [(A_{ij} - \overline{A}) u u^T (A_{ij} - \overline{A})^T]. & (16) \end{aligned}$$

Then, we compute $S_T(u)'s$ eigenvectors $v_1, \cdots, v_\delta$ corresponding to the δ largest eigenvalues; and set $V := [v_1 \ \cdots \ v_\delta]$, $\overline{S}_b(u) := V^T S_b(u) V$, and $\overline{S}_w(u) := V^T S_w(u) V$. Furthermore, we compute the generalized eigenvectors $y_1, \cdots, y_\gamma$ of $\overline{S}_b(u)$ and $\overline{S}_w(u)$ corresponding to the γ largest eigenvalues; and set $w_1 := V y_1, \cdots, w_\gamma := V y_\gamma$. Then, we obtain $W^* = [w_1 \ w_2 \ \cdots \ w_\gamma]$.

Case 2. Suppose that W is fixed. In this case, we know from Theorem 4.1 that the constraint in (13) means that $\mathrm{tr}(W^T S_w(u) W) = \gamma$, where γ is the number of projection basis vectors. Furthermore, we have

$$\mathrm{tr}(W^T S_w(u) W) = u^T L_w(W) u = \gamma, \quad \text{i.e.,} \quad u^T[(1/\gamma) L_w(W)]u = 1.$$

Therefore, if W is fixed, the optimization problem (13) is equivalent to

$$\max_u \ u^T L_b(W) u \quad \text{s.t.} \quad u^T[(1/\gamma) L_w(W)]u = 1,$$

which implies that we only need to find the generalized eigenvector of matrices $L_b(W)$ and $(1/\gamma) L_w(W)$ corresponding to the largest eigenvalue.

By combining **Case 1** with **Case 2**, we now design an alternating direction method to solve the model (13), which is described in Algorithm 4.1.

Algorithm 4.1. *(**TWSBF+LDA**)*

1: *Let $L_{ij}, H_{ij}, V_{ij}, D_{ij}$ ($i = 1, 2, \cdots, c$, $j = 1, 2, \cdots, m_i$) be four wavelet sub-bands obtained by the t-level 2D-DWT.*

2: *Implement LDA dimension reduction on L_{ij} and obtain DL_{ij}.*

3: *Denote $A_{ij} = [H_{ij} \ V_{ij} \ D_{ij}]$.Initialize $u = u^{[0]}$ and set $P_i = m_i/m$ ($i = 1, 2, \cdots, c$). Set $k = 0$.*

4: *Compute $S_b(u)$ and $S_w(u)$ by (10) and (11), respectively.*

5: *Compute $S_T(u)$ by (16); and its eigenvectors $v_1, \cdots, v_\delta$ corresponding to the δ largest eigenvalues of $S_T(u)$.*

6: *Let $V = (v_1, \cdots, v_\delta)$. Compute $\overline{S}_b(u)$ and $\overline{S}_w(u)$ according to $\overline{S}_b(u) = V^T S_b(u) V$, $\overline{S}_w(u) = V^T S_w(u) V$, and their generalized eigenvectors $y_1, \cdots, y_\gamma$ corresponding to the γ largest eigenvalues of $\overline{S}_b(u)$ and $\overline{S}_w(u)$. Let $w_1 = V y_1, \cdots, w_\gamma = V y_\gamma$, $W^{[k+1]} = [w_1 \ w_2 \ \cdots \ w_\gamma]$.*

7: *Compute* $L_b(W^{[k+1]})$ *and* $L_w(W^{[k+1]})$ *by (14) and (15), respectively; and their generalized eigenvector* $u^{[k+1]}$ *corresponding to the largest eigenvalue of* $L_b(W^{[k+1]})$ *and* $L_w(W^{[k+1]})$*.*

8: *If* $k > 0$ *and*

$$\left|J(W^{[k+1]}, u^{[k+1]}) - J(W^{[k]}, u^{[k]})\right| < \varepsilon,$$

then

Let $u^* = u^{[k+1]}$, $W^* = W^{[k+1]}$. *Stop.*

else

Update $u = u^{[k+1]}$, $k = k + 1$. *Go to 3.*

end if

9: *Compute* DS_{ij} *through* u^* *and* W^*. *Then implement feature-level fusion on* DL_{ij} *and* DS_{ij} *and obtain* F_{ij}.

5. Numerical Experiments

In this section, we first compare our methods: TWSBF + PCA (i.e., Algorithm 3.1) and TWSBF + LDA (i.e., Algorithm 4.1) with the two classical wavelet-based methods:

- DWT + PCA [25], i.e., an image of face is transformed by using the t-level wavelet transform, and the dimension of the final obtained low frequency sub-band is reduced by using the classical PCA;
- DWT + LDA [11], i.e., an image of face is transformed by using the t-level wavelet transform, and the dimension of the final obtained low frequency sub-band is reduced by using the strategy of the classical PCA plus the classical LDA;

Then, we compare our methods with some popular methods:

- GMTR [26], i.e., an image of face is divided into appropriate sub-regions and then transformed by Gabor transform, and the Gabor magnitude is selected as texture representation. The dimension is reduced by using the classical PCA.

- PCRC [27], i.e., patch based collaborative representation classification.
- PNN [28], i.e., patch based nearest neighbor classification.
- SRC [29], i.e., sparse representation based classification.
- CRC [30], i.e., collaborative representation based classification.

We implement these methods on the FERET, ORL and AR face databases.

A basic face recognition system is composed of two stages: training process and recognition process; and our methods focus on the first stage. However, recognition process is also very important; and in this stage, classifier plays a vital role. There are many classifiers such as nearest neighbor classifier (NN) [31], nearest subspace classifier [32], support vector machine (SVM) [33], and so on. In our experiments, we use NN and SVM. NN is the most basic and popular classifier. As an evolution of SVM, k-SVM completes the transformation from the original feature space to a higher dimension space, which ensures the better discrimination of face features.

Our experiments are made on a PC platform with 64-bit win8 operating system, Intel Core i5-3550S CPU, and 8G memory.

5.1. Experiments on the FERET Database

The FERET image corpus was assembled to support government monitored testing and evaluation of face recognition algorithms using standardized tests and procedures. The final corpus, consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles. In our experiments, FERET database contains 1400 gray level images of 200 individuals (each person has 7 different images) and each image is manually cropped and resized to 80×80 pixels.

Firstly, we compare our methods (i.e., TWSBF + PCA and TWSBF + LDA) with two wavelet based methods (i.e., DWT + PCA and DWT + LDA), where NN is used to classify the query sample. Specifically, each original image is transformed through the 4-level wavelet transform using the haar wavelet. We select randomly i samples per person for training and the remaining samples for testing, where $i = 2, 3, 4, 5, 6$, respectively. We repeat the experiments 10 times with each of the above four methods, and calculate the average recognition rate for every case. The numerical results are shown in Table 1.

Table 1: Average Recognition Rates on the FERET Database

Method	2	3	4	5	6
DWT+PCA	34.09%	34.96%	40.56%	42.75%	54.00%
TWSBF +PCA	**38.37%**	**39.65%**	**45.51%**	**47.17%**	**59.55%**
DWT+LDA	41.78%	43.40%	45.93%	46.60%	58.90%
TWSBF +LDA	**50.73%**	**54.06%**	**57.26%**	**59.45%**	**73.00%**

From Table 1, it is easy to see that the proposed methods show higher recognition rates than the basic DWT + PCA and DWT + LDA methods no matter how many training samples per person are used. This suggests that our fusion technique is effective and the high frequency sub-bands indeed play a great role in face recognition.

Secondly, we compare our methods with other popular methods. We select randomly i samples per person for training and the remaining samples for testing, where $i = 2, 3, 4, 5, 6$, respectively; and then, repeat the experiments 10 times and calculate the average recognition rates. In the experiments, the settings of other methods' experimental parameters are as follows:

- PNN: the size of the square patches is set as 10.
- PCRC: the regularization parameter is set as 0.001, the size of the square patches is set as 10, the overlap parameter is set as 5, the neighborhood size is set 0.
- SRC: the regularization parameter is set as 0.001.
- CRC: the regularization parameter is set as 0.005.

The average recognition rates are listed in Table 2.

From Table 2, we can see that our methods are better than other popular methods in most cases, especially TWSBF + LDA, because LDA makes full use of the category information compared with PCA.

5.2. Experiments on the ORL Database

The ORL database contains images of 40 individuals, and each people has 10 different images with the size of 112 × 92 pixels. For some persons, the images were taken at different times, varying the lighting, facial expressions (open/close eye, smiling/not smiling) and facial details (glasses/no glasses).

Table 2: Average Recognition Rates on the FERET Database

Method	2	3	4	5	6
GMTR	6.21%	10.70%	13.88%	14.72%	20.25%
PCRC	22.85%	27.65%	29.00%	23.37%	24.60%
CRC	28.75%	35.57%	40.11%	39.30%	48.55%
SRC	28.57%	32.23%	35.93%	33.07%	40.25%
PNN	39.92%	44.81%	53.96%	56.87%	**74.30%**
TWSBF +PCA	38.37%	39.65%	45.51%	47.17%	59.55%
TWSBF +LDA	**50.73%**	**54.06%**	**57.26%**	**59.45%**	73.00%

All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

Firstly, we compare our methods (i.e., TWSBF + PCA and TWSBF + LDA) with two wavelet based methods (i.e., DWT + PCA and DWT + LDA). Specifically, each original image is transformed through the 4-level wavelet transform using the haar wavelet. We select randomly i samples per person for training and the remaining samples for testing, where $i = 3, 4, 5, 6, 7, 8$, respectively. We repeat the experiments 10 times with each of the above four methods, and calculate the average recognition rate for every case. We use NN to classify the query sample. The numerical results are shown in Table 3.

Table 3: Average Recognition Rates on the ORL Database

Method	3	4	5	6	7	8
DWT+PCA	86.64%	90.54%	92.60%	94.06%	95.83%	96.00%
TWSBF +PCA	**88.10%**	**92.58%**	**94.25%**	**95.50%**	**97.25%**	**97.50%**
DWT+LDA	91.00%	94.12%	97.10%	97.62%	97.66%	98.25%
TWSBF +LDA	**92.07%**	**95.04%**	**97.15%**	**97.87%**	**98.66%**	**99.00%**

From Table 3, it is easy to see that the proposed methods show higher recognition rates than the basic DWT + PCA and DWT + LDA methods.

Secondly, we compare our methods with other popular methods. Similarly, we select randomly i samples per person for training and the remaining samples for testing, where $i = 3, 4, 5, 6, 7, 8$, respectively. We repeat the experiments 10 times with each of the above four methods, and calculate the average recognition rate for every case. Experimental parameter settings are as follows:

- PNN: the size of the square patches is set as 10.
- PCRC: the regularization parameter is set as 0.001, the size of the square patches is set as 10, the overlap parameter is set as 5, the neighborhood size is set 0.
- SRC: the regularization parameter is set as 0.001.
- CRC: the regularization parameter is set as 0.005.

The average recognition rates are listed in Table 4.

Table 4: Average Recognition Rates on the ORL Database

Method	3	4	5	6	7	8
GMTR	76.53%	83.45%	87.35%	89.62%	91.91%	91.37%
PCRC	67.64%	72.79%	76.45%	77.93%	79.58%	79.62%
CRC	86.46%	89.20%	92.25%	92.87%	93.00%	93.37%
SRC	87.28%	89.91%	92.30%	93.00%	93.25%	93.75%
PNN	88.25%	92.66%	94.95%	96.12%	96.66%	97.25%
TWSBF +PCA	88.10%	92.58%	94.25%	95.50%	97.25%	97.50%
TWSBF +LDA	**92.07%**	**95.04%**	**97.15%**	**97.87%**	**98.66%**	**99.00%**

From Table 4, it is easy to see that TWSBF + LDA achieves the higher recognition rates on all experiments; and TWSBF + PCA has also good performances compared with other popular methods.

5.3. Experiments on the AR Database

The AR database contains over 4,000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The images of most persons were taken in two sessions (separated by two weeks). Each section contains 13 color images and 120 individuals (65 men and 55 women) participated in both sessions. In our experiments, each image is manually cropped and resized to 50×40 pixels.

Firstly, each original image is transformed through 2-level wavelet transform using the 4-th order Daubechies wavelet because images in AR Database are relatively small. For comparison purpose, the methods of TWSBF + P-CA, TWSBF + LDA, DWT + PCA, and DWT + LDA are implemented, respectively. We select randomly i samples per person for training and the remaining samples for testing, where $i = 3, 4, 5, 6, 7, 8$, respectively. We repeat the experiments 10 times with each of the above four methods, and calculate the average recognition rate for every case. We use k-SVM to classify and the parameters of k-SVM are set as follows: the penalty parameter is set as 128 and the parameter given in Gaussian radial basis function is set as 0.0078425. The numerical results are shown in Table 5.

Table 5: Average Recognition Rates on the AR Database

Method	3	4	5	6	7	8
DWT+PCA	58.30%	65.70%	68.96%	75.69%	80.71%	83.06%
TWSBF +PCA	**61.90%**	**69.99%**	**73.78%**	**79.50%**	**83.99%**	**86.25%**
DWT+LDA	60.85%	68.22%	73.30%	79.58%	83.92%	87.32%
TWSBF +LDA	**65.48%**	**73.47%**	**81.22%**	**84.89%**	**88.51%**	**90.87%**

From Table 5, it can be easily seen that our methods achieve the higher recognition rates on all experiments than DWT + PCA and DWT + LDA. And as Table 5 shows, though the number of training images changes, the recognition rates of the proposed methods are always higher than those of the DWT + PCA and DWT + LDA methods, which indicates that the proposed methods are more robust with the number of training samples varying.

Secondly, we compare our methods with other popular methods. We select randomly thirteen training images per person and the remaining are for testing; repeat the experiments 10 times, and then calculate the average recognition rates. Experimental parameter Settings are as follows:

- PNN: the size of the square patches is set as 10.
- PCRC: the regularization parameter is set as 0.001, the size of the square patches is set as 10, the overlap parameter is set as 5, the neighborhood size is set 0.
- SRC: the regularization parameter is set as 0.001.
- CRC: the regularization parameter is set as 0.005.

The average recognition rates are listed in Figure 4.

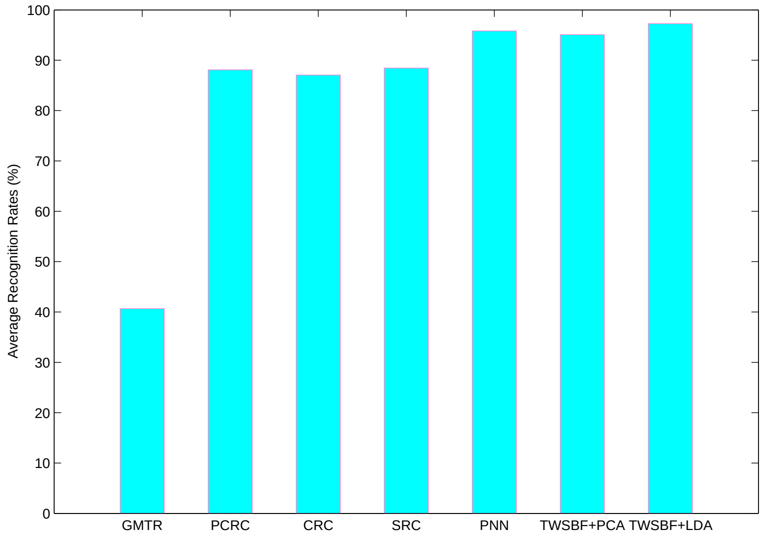


Figure 4: The Average Recognition Rates on AR Database

From Figure 4, we can see that our method of TWSBF + LDA is superior to the other popular methods and TWSBF + PCA is better or at least not worse than the popular methods.

6. Remarks

In this paper, we proposed two effective methods of feature extraction for face recognition. There are four important points in our methods: Firstly, the data fusion technique proposed in this paper ensures four top-level's wavelet sub-bands are all used, which avoids the top-level's information to be lost. Secondly, in order to make the proposed high frequency sub-bands pixel-level fusion technique more effective, by combining it with PCA and LDA respectively, we converted the problem of finding the best fusion coefficients to two optimization problems and designed two alternative direction methods to solve the corresponding problems. Thirdly, the optimal fusion coefficients of wavelet sub-bands and the optimal transformation matrices of dimension reduction were obtained simultaneously. Fourthly, we combined feature-level fusion with pixel-level fusion effectively. Experimental results on the FERET, ORL and AR face databases demonstrate the effectiveness and robustness of the proposed methods for face recognition.

Some further issues are worth studying in the future. Firstly, our methods were proposed for face recognition. We believe the methods proposed in this paper can be further used in some other fields about data fusion of images. Secondly, our methods were proposed by combining the sub-band fusion technique with PCA and LDA, respectively. In fact, many other approaches for dimension reduction have been proposed in the literature, such as independent component analysis, kernel principal component analysis, kernel linear discriminant analysis, and so on. It is possible that some more effective methods can be proposed by combining the sub-band fusion technique proposed in this paper with other approaches for dimension reduction. Thirdly, in our methods, only four top-level's wavelet sub-bands were effectively fused. It is possible that the proposed fusion technique of the top-level's sub-bands can be further developed to fuse the information of more wavelet sub-bands including those located in other wavelet levels instead of the top-level only.

Acknowledgments

This work was partially supported by the National Natural Science Foundations of China (Grant No. 11171252).

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, Face recognition: a literature survey, ACM Computing Surveys, 35(4)(2003) 399-458.

[2] Y. Cheng, Y.K. Hou, C.X. Zhao, Z.Y. Li, Y. Hu, C.L. Wang, Robust face recognition based on illumination invariant in nonsubsampled contourlet transform domain, Neurocomputing, 73(2010) 2217-2224.

[3] X. Zhang, Y. Gao, Face recognition across pose: a review, Pattern Recog., 42(2009) 2876-2896.

[4] A. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intell., 24(6)(2002) 748-763.

[5] J. Zhao, Z. Zhou, F. Cao, Human face recognition based on ensemble of polyharmonic extreme learning machine, Neural Comput. Appl., (2013). doi:10.1007/s00521-013-1356-4

[6] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 1986.

[7] V. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell., 19(7)(1997) 711-720.

[8] K.C. Kwak, W. Pedrycz, Face recognition using fuzzy integral and wavelet decomposition method, IEEE Trans. Syst., Man, Cybern, 34(1)(2004) 1666-1675.

[9] D. Li, W. Pedrycz, N.J. Pizzi, Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification, IEEE Trans. Biomed. Eng., 52(6)(2005) 1132-1139.

[10] C.L. Fan, X.T. Chen, N.D. Jin, Research of face recognition based on wavelet transform and principal component analysis, in: ICNC, 2012, pp. 575-578.

[11] J.T. Chien, C.C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, IEEE Trans. Pattern Anal. Mach. Intell., 24(12)(2002) 1644-1649.

[12] V. Vidya, N. Farheen, K. Manikantan, S. Ramachandran, Face recognition using threshold based DWT feature extraction and selective illumination enhancement technique, in: ICCCS, 2012, pp. 334-343.

[13] D.V. Jadhav, R.S. Holambe, Feature extraction using Radon and wavelet transforms with application to face recognition, Neurocomputing, 72(2009) 1951-1959.

[14] B. Li, Y.H. Liu, When eigenfaces are combined with wavelets, Knowledge-Based Systems, 15(2002) 343-347.

[15] H.F. Hu, Variable lighting face recognition using discrete wavelet transform, Pattern Recognit. Lett., 32(13)(2011) 1526-1534.

[16] J. Yang, J.Y. Yang, D. Zhang, J.F. Lu, Feature fusion: parallel strategy vs. serial strategy, Pattern Recognit., 36(6)(2003) 1369-1381.

[17] M. Kumar, S. Dass, A total variation-based algorithm for pixel-level image fusion, IEEE Trans. Image Process., 18(9)(2009) 2137-2143.

[18] G. Pajares, J.M. Cruz, A wavelet-based image fusion tutorial, Pattern Recognit., 37(2004) 1855-1872.

[19] D.K. Sahu, M.P. Parsai, Different image fusion techniques - a critical review, IJMER., 2(5)(2012) 4298-4301.

[20] N. Mitianoudis, T. Stathaki, Pixel-based and region-based image fusion schemes using ICA bases, Info. Fusion, 8(2007) 131-142.

[21] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.

[22] P. Lancaster, M. Tismenetsky, The Theory of Matrices, 2nd ed, Academic Press, Orlando, 1985.

[23] C.R. Rao, The utilization of multiple measurements in problems of biological classification. J. R. Stat. Soc. Ser. B, 10(1948) 159-203.

[24] C.R. Rao, Linear Statistical Inference and its Application, 2nd ed, New York, 2002.

[25] M. Safari, M.T. Harandi, B.N. Araabi, A SVM-based method for face recognition using a wavelet PCA representation of faces, in: ICIP, 2004, pp. 853-856.

[26] L. Yu, Z.S. He, Q. Cao, Gabor texture representation method for face recognition using the Gamma and generalized Gaussian models, Image Vision Comput., 28(2010) 177-187.

[27] P.F. Zhu, L. Zhang, Simon C.K. Shiu, Multi-scale patch based collaborative representation for face recognition with margin distribution optimization, in: ECCV, 2012, pp. 822-835.

[28] R. Kumar, A. Banerjee, B.C. Vemuri, H. Pfister, Maximizing all margins: Pushing face recognition with kernel plurality, in: ICCV, 2011, pp. 2375-2382.

[29] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell., 31(2)(2009), 210-227.

[30] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition? in: ICCV, 2011, pp. 471-478.

[31] R. Duda, P. Hart, D. Stork, Pattern Classification, 2nd ed, John Wiley Sons, 2001.

[32] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman, Clustering appearances of objects under varying illumination conditions, in: CVPR, 2003, pp. 11-18.

[33] V.N. Vapnik, Statistical Learning Theory, Wiley-Interscience, 1998.

[34] J.C. Platt, N. Cristianini, J.S. Taylor, Large margin DAGs for multiclass classification, in: NIPS, 2000, pp. 547-553.