International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.6, No.5 (2013), pp.423-436
http://dx.doi.org/10.14257/ijsip.2013.6.5.37

Face Recognition Based on Multi-classifierWeighted Optimization and Sparse Representation

Deng Nan[1], Zhengguang Xu[2] and ShengQin Bian[3]
[1,2,3]*Institute of control science and engineering,*
University of Science and Technology Beijing
[1,2,3]*30 Xueyuan Road, Haidian District, Beijing 100083 P. R.China*
[1]*dengnan666666@163.com,* [2]*xzg_1@263.net,* [3]*xiaobian@ustb.edu.cn*

Abstract

Facial recognition (FR) is a challenging area of research due to difficulties with robust FR when the number of training samples is very small. The state-of-the-art sparse representation-based classification (SRC) shows very excellent FR performance. However, the recognition rate of SRC will drop dramatically when the number of training samples per class is very limited. To solve these issues, we propose a weighted multi-classifier optimization and sparse representation based (WMSRC) method for FR, which efficiently combines the local and global characteristics of face images. A face image is firstly divided into continuous but non-overlapped blocks by multi-resolution based blocking and each block is sparsely represented over the corresponding set of blocks of all training samples. The multi-scale SRC classifiers are then established and associated with different weights based on sub-block dictionary learning. According to the multiple voting results of the classifiers, the weights of multi-classifiers are optimized by a least-squares optimization equation with l_2 *-norm regularization. Finally, the classification results of all the blocks are combined by a weighted fusion criterion. Our experiments show that the WMSRC algorithm outperforms many existing block-based sparse representation classification algorithms, especially for FR when the available training samples per subject are very limited.*

Keywords: *Face recognition, Sparse representation, Dictionary learning, Multi-classifier Weight Optimization, Multi-classifier fusion*

1. Introduction

Face recognition (FR) [1] has been increasingly applied in video surveillance, ID authentication, information security and other fields in recent years. With respect to a sample library of a certain scale, certification cooperation and ideal imaging conditions, the correct recognition rate of frontal face recognition, which has been applied in practice, can reach up to more than 90%. However, factors such as illumination, posture, facial expression, occlusion and so on will affect the robustness of FR, wherein, the small sample size (SSS) problem is an important factor [2]. For example, on some special occasions such as passport verification, law enforcement and ID authentication, one can only get very few training samples, and sometimes even only one single sample, thus, the number of dimensions of image samples is much larger than the number of training samples. This will result in the failure of the distance measurement in pattern recognition, thus quickly decreasing or depriving of the recognition performance of traditional PCA, LDA, LPP face recognition methods and their derivative ones [3].

Sparse representation face recognition (SRC) [4] is modeled based on the image subspace assumption [5], which means that various training images can form an independent partition for the image subspace, and any test image can be expressed by the linear combination of the same face training image. Through sparse representation, faces

ISSN: 2005-4254 IJSIP

to be tested are approximately expressed as a linear sparse combination of all types of training faces, and testing samples for the minimum reconstruction residue of various types of training samples are determined by calculating the sparse combination coefficient. This method has been proven not only to be able to effectively improve recognition performance, but also can effectively reduce the effect of occlusion recognition. However, when the training sample size is very small, it is difficult for test samples to be adequately and approximately represented by all training faces, thus significantly reducing recognition performance.

The SSS problem [6-8] is generally solved by generating virtual samples and introducing training sets with the following three methods: the first method [9,10] is based on local features, generally including the partitioning of faces, the extraction of local features, formation of local classifiers and combination of all local classifiers for judgment of categories in the final stage; in this method, the size of the face block will affect the recognition results, and partitioning is generally based on experience; the second method [11, 12] recognizes faces by combining local and global features of faces based on the complementation of local features and global features of faces; the third one [13, 14] extracts different types of facial features for classifier fusion recognition through several facial feature extraction methods. For example, the commonly used LBP feature, Gabor feature [15] and SIFT feature are recognized using this method.

Multi-scale phenomenon is extremely common in the real world. As face images will present different structures and features at different scales, the characteristics of images extracted based on multi-scale [16, 17] will not only give more effective face information, but also lead to simple and effective models insensitive to the scale through the combination of information at different scales, thus improving recognition performance.

Classifier fusion can be roughly divided into abstraction, arrangement, and fraction fusions based on different classifier outputs [18]. Fraction fusion is one of the most commonly-used methods in classifier fusion because it can not only shield the feature diversity of the images and the complexity of the process of the identification, but also store the similarity metrics of various features. Those methods based on fraction fusion can be divided into two categories. One category of the methods is based on the fixed rules such as product, sum, maximum, minimum, mid-value, and voting rules. Another different category of the methods is based on training rules such as support vector machine (SVM) meta classifier and weight classifier. Of them, the weight classifier is characterized by simple algorithm, flexible use and wide application.

Based on the above, to improve the performance of SRC in the SSS problem, we proposed a multi-classifier weighted optimization and sparse representation based (WMSRC) method for face recognition with limited samples. This method combines the multi-resolution based blocking and SRC to obtain local and overall characteristics of face image. According to the multiple voting results of the classifiers, the classification results of all the blocks are combined by a weighted fusion criterion after the weights of multi-classifiers are optimized by a least-squares optimization equation with l_2-norm regularization and is able to achieve a higher recognition rate compared to the existing SRC and BSRC method [4, 16, 17]. Pengfei Zhu1 *et al.*, [17] proposed a multi-scale Patch based Collaborative Representation for Face Recognition with Margin Distribution Optimization sparse representation model for face recognition. However, the essence and the way of multi-resolution based blocking of this method are different from the method proposed in this paper.

The main contributions of this paper are as follows: First, the SRC method containing overall characteristic information was introduced to conduct local blocking of the face image, which excavated the local and overall characteristics of the face image for FR. Second, the sub-blocks of the training samples are selected to constitute the original training set. The simpler and more compactly supported dictionary base of each sub-block is obtained through the K-SVD dictionary learning algorithm [19]. Thirdly, a least-squares optimization equation for the mean square error containing weight vector parameters and constraint conditions [17] is established according to the multiple voting results of the classifiers, and the optimal weight vector of each classifier is solved. Finally, the weighted fusion criterion is used for recognition of fraction fusion [20].

2. Sparse Representation-based Classification for FR

The set of training samples of the i_{th} class can be expressed as $\Phi_i = [A_{i,1}, \cdots, A_{i,j}, \cdots A_{i,n_i}] \in R^{m \times n_i}$, where $A_{i,j}, j = 1, \cdots, n_i$ is a m dimensional column vector composed of the j_{th} sample image of the i_{th} training class. In the sparse representation model, $A_{i,j}$ is also called atom, and the combination of the atoms composes a dictionary. The testing sample $x \in R^m$ from this class can be represented by $x = \sum_{j=1}^{n_i} \theta_{i,j} A_{i,j} = \Phi_i \theta_i$, where $\theta_i = [\theta_{i,1}, \theta_{i,2}, \cdots, \theta_{i,n_i}] \in R^{n_i}$ are the weights. Denoting the over-complete dictionary matrix composed by atoms from the whole training images of all k object classes as by $\Phi = [\Phi_1, \ldots \Phi_i \ldots \Phi_k] \in R^{m \times n}$, the testing image of the i_{th} training subject can be represented by $x = \Phi\theta \in R^m$,where $\theta = [\theta_1; \cdots; \theta_i; \cdots; \theta_k] = [0, \cdots, 0, \theta_{i,1}, \cdots \theta_{i,n_i}, 0, \cdots, 0]^T \in R^m$.All image samples are normalized and their dimensions are reduced as $\tilde{x} = T\Phi\theta = \tilde{\Phi}\theta \in R^d$,where $T \in R^{d \times m}, d \le m$ is the feature transformation matrix.

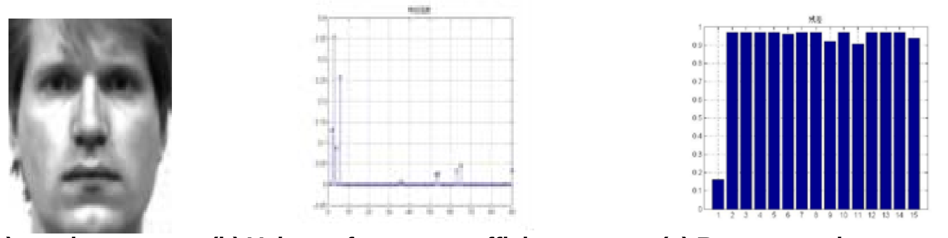
Obviously the sparser θ is, the more concise the image can be represented. The sparsest representation of image can be acquired by solving the l_0-minimization problem of θ , i.e. $\hat{\theta}_0 = \arg\min \|\theta\|_0, s.t. \|\tilde{x} - \tilde{\Phi}\theta\| \le \varepsilon$, where $\|\cdot\|_0$ denotes l_0 -norm . In practice, this problem is efficiently solved by minimizing l_1-norm [11, 12]:

$$\hat{\theta}_1 = \arg\min_{\theta} \left\{ \left\| \tilde{x} - \tilde{\Phi}\theta \right\|_2^2 + \lambda \|\theta\|_1 \right\} \tag{1}$$

wherein λ is regularization parameter. To identify the object, the reconstruction errors are calculated by $r_i(\tilde{x}) = \|\tilde{x} - \tilde{\Phi}\delta_i(\hat{\theta}_1)\|_2 (i = 1, \cdots, k)$, where $\delta_i(\theta) \in R^n$ is the characteristic function which selects the coefficients of the i_{th} class. And the classification is based on the minimization reconstruction errors:

$$identify(\tilde{x}) = \arg\min r_i(\tilde{x}) \tag{2}$$

The recognition of a testing image by SRC is shown in Figure 1.



(a) test image **(b) Values of sparse coefficients** **(c) Reconstruction errors**

Figure 1. Recognition of a Testing Image by SRC

3. Sparse Representation based Multi-Classifier Weighted Optimization and Fusion for FR

We first introduce the overall process of the multi-classifier weighted optimization and sparse representation based (WMSRC) method proposed in this paper. Then, the key points of the method are presented.

3.1. Total Flow of WMSRC Method

During the training stage, the training samples are processed by multi-resolution blocking, and the non-lap sub-blocks of each training image with different resolutions and space positions are obtained. Then, the characteristic transformation matrix is calculated to reduce the dimension of sub-blocks by principal component analysis (PCA), linear discriminant analysis (LDA), down-sampling, and the linear projection generated by a Gaussian random matrix method (Random) [21] separately.

During the testing stage, the testing image is first processed by multi-resolution blocking to obtain sub-blocks as introduced above. The dimension of each sub-block is reduced by the characteristic transformation matrix obtained in the training stage. Then, the sub-blocks of the training samples are selected to constitute the original training set and the simpler and more compactly supported dictionary base of each sub-block is obtained through the K-SVD dictionary learning algorithm and greedy tracking algorithm. After that, multi-scale SRC classifiers are established and associated with different weights. Next, a least-squares optimization equation for the mean square error containing weight vector parameters and constraint conditions is established according to the multiple voting results of the classifiers, and the optimal weight vector of each classifier is solved. Finally, the weighted fusion criterion is used for recognition. The system diagram of the proposed method is shown in Figure 2.

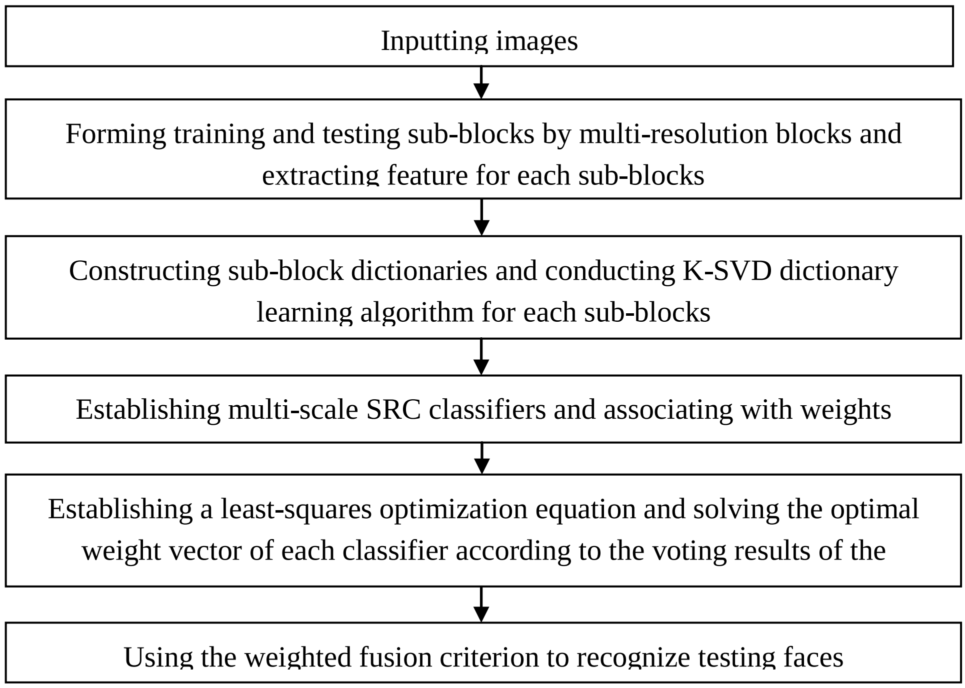


Figure 2. System Diagram of the Proposed WMSRC-based FR Method

3.2. Multi-resolution Block and Dimensionality Reduction of Face Image

Multi-resolution decomposition of face image contains the image information both locally and globally. When face images with different resolutions are obtained by the multi-resolution decomposition, images with higher resolution can be divided into more blocks due to the fact that human eyes are able to observe more local information of object with higher resolution. Accordingly, images with lower resolution can be divided into fewer blocks. The adaptive blocking refers to the number of the sub-blocks depends on the resolution of the image in partitioning, which accords with human vision.

We first evenly divided the normalized image into sixteen blocks by image sub-sampling [4], and obtained the first layer of sub-blocks. We conducted down-sampling to evenly divide the sub-blocks into four blocks to obtain the second layer of sub-blocks; and continued down-sampling to obtain the third layer of sub-blocks. In total, twenty-four non-overlapping image sub-blocks were obtained as shown in Figure 3. Finally, after partition, the PCA, LDA, down-sampling, and Random methods are utilized in this paper to conduct feature extraction and dimension reduction on training and testing datasets.

(a) The first layer of sub-blocks (b) The second layer of sub-blocks (c) The third layer

Figure 3. Multi-resolution Block

3.3. K-SVD-based dictionary learning of sub-blocks

When selecting dictionaries, it is necessary to select a great number of training samples for each person if the original training sample matrices are directly used for building a dictionary, so as to improve the representation ability of the dictionary, thus making it more complex to solve sparse problems. At the same time, it is necessary to select training samples carefully to build the dictionary so that the dictionary atoms will cover the sub-space composed of various samples as reasonably as possible. For example, generally, the training samples with different gestures and different expressions, under different illumination conditions and without manual occlusion are selected to build a dictionary. In contrast, dictionary learning will result in dictionaries which are sparser than the given training sample set or the built original dictionary so that fewer atoms will be able to represent the new sample of the same category, and sparser and more accurate sparse solutions will be obtained.

K-SVD dictionary learning is used to find the optimal basis under sparse representation to meet its unique conditions. It can effectively reduce the number of atoms, improve atomic irrelevance, and obtain more sparse and accurate sparse solution.

Assuming that $X=[x_1,x_2,\cdots,x_M]$ is a training set, D is an over-complete dictionary, dictionary dimension is n, K is the number of atoms in the dictionary, M is the number of samples, and $A=[\alpha_1,\alpha_2,\cdots,\alpha_M]$ is the corresponding coefficient of characterization, the sparse representation model is $\hat{\alpha}=\arg\min_{\alpha}\{\|X-D\alpha\|_2^2+\lambda\|\alpha\|_0\}$. And the iterative solution can be conducted in the following two stages.

The first stage is sparse coding. Assuming that dictionary D (first K samples can be used as an initialized dictionary) is a constant value, the sparse representation vector α_i of each input training sample can be calculated by $\boldsymbol{\alpha}_i=\min\limits_{\boldsymbol{\alpha}}|\boldsymbol{\alpha}|_0, \quad s.t. \quad \|\mathbf{x}_i-D\boldsymbol{\alpha}\|\le\varepsilon;\ i=1,2,\ldots,M$.

The second stage is dictionary updating strategies are similar to that in the sparse coding stage, *i.e.*, first fix the sparse representation vector matrix A obtained in sparse coding stage, assuming that α^k is the k_{th} column of A, successively upgrade the first row of the dictionary (every atom d_k), and to $E=\left\|(X-\sum_{j\ne k}\mathbf{d}_j\boldsymbol{\alpha}^j)-\mathbf{d}_k\boldsymbol{\alpha}^k\right\|_F^2=\|E_k-\mathbf{d}_k\boldsymbol{\alpha}^k\|_F^2$

In this formula, E_k represents the error matrix after removing d_k from dictionary D. Assuming that $\omega_k=\{i|1<i<K,\alpha^K(i)\ne0\}$ is a non-zero index entry of α^k, defining the matrix Ω_k with a size of $M\times|\omega_k|$, setting it to be 1 at $(\omega_k(i),i)$ and 0 at other places, we can obtain dimensionality reduction matrix $\alpha_R^k=\alpha^k\Omega_k$ and $\alpha_R^k\in1\times|\omega_k|$ by multiplying α^k and Ω_k,

i.e., remove non-zero item. In a similar way, we can obtain $x_R^k = x\Omega_k$, $x_R^k \in n \times |\omega_k|$ and $E_R^k = E_k\Omega_k$. After multiplying the above formula by Ω_k, we can gain the formula $\|E_k\Omega_k - d_k\alpha^k\Omega_k\|_F^2 = \|E_R^k - d_k\alpha_R^k\|_F^2$. Finally, we can conduct singular value decomposition (SVD) for E_R^k with the formula $E_R^k = U\Delta V^T$, upgrade d_k using the first column of matrix U after decomposition, upgrade α_R^k using the first column of $V \times \Delta(1,1)$, and upgrade the first column of dictionary sequentially until the end.

The commonly used algorithms for l_0-norm minimization include greedy tracking [8] and relaxation optimization [1]. This paper conducted sparse decomposition by the orthogonal matching pursuit (OMP) method [8].

3.4. Establishment of Multi-classifier

We conducted multi-resolution blocks for the image, and established the sparse characterization model of multi-classifiers by $\hat{\rho}_j = \arg\min_{\rho_j}\{\|y_j - M_j\rho_j\|_2^2 + \lambda\|\rho_j\|_0\}$. In $M_j = [M_{j1}, M_{j2}, \cdots, M_{jc}]$, M_j represents the classifier dictionary of the j_{th} sub-block, and M_{jk} represents the classifier dictionary of the j_{th} sub-block in the k_{th} category of Class C.

We solved each classifier dictionary and sparse characterization vector by K-SVD and OMP, and solved the classifier recognition results z_j using minimum reconstruction residual as $z_j = identify(y_j) = \arg\min_k(r_{jk})$.

3.5. Multi-classifier Weighted Optimization

Different classifiers contribute differently to the recognition. That is, they have different identification abilities. During fusion, the classifiers should be assigned weights according to their identification abilities. The results of multiple votes reflect the degree of the judgment and recognition [17]. So the weights can be calculated according to the results of the multiple votes for the classifiers. If the results of multiple votes and the weight vectors are used to compose an optimal equation with regularized limitations, the obtained weight vectors of the classifiers will inevitably be more stable.

Based on the above ideas, test sample containing multiple classes is defined as $\{g_i\}, i = 1,2,\cdots c$, where c is the total number of classes. $s = 21$ is the total number of blocks. $w_j(j = 1,2,\cdots s)$ represents the weights of the classifiers, satisfying $\sum_{j=1}^{s} w_j = 1$, d_{ij} stands for the classification result of the j_{th} classifier in the i_{th} class. When the recognition result of the classifier is correct, $d_{ij} = +1$; otherwise, that is, when the recognition is wrong, $d_{ij} = -1$. When the recognition system refuses to recognize, d_{ij} approaches negative infinity. The class function of the test sample is defined as $f(g_i) = \sum_{j=1}^{s} w_j d_{ij}$ which can reflect the degree of wrong classification. Obviously, the value of $f(g_i)$ ranges from -1 to +1. The greater the value, the closer the tested face is to a certain class. Then the loss function for determining a class is $l_{x_i} = l(f(g_i)) = l(\sum_{j=1}^{s} w_j d_{ij})$. The overall mean square error of the tested face is defined as $l(S) = \sum_{i=1}^{c} l_{g_i} = \sum_{i=1}^{c}[1 - f(g_i)]^2 = \sum_{i=1}^{c}[1 - \sum_{j=1}^{s} w_j d_{ij}]^2 = \|a - Dw\|^2$, wherein a is a vector with all elements being 1 and the length is s; D is the binary classification result of each sub-block. The least-squares optimization equation for the mean square error with W parameters and l_2 norm is defined as follows:

$$\hat{w} = \arg\min_w(\|a - Dw\|_2^2 + \tau\|w\|_2), s.t. \sum_{j=1}^{S} w_j = 1, w_j > 0, j = 1,\cdots,s \tag{3}$$

wherein τ is regularization parameter. In the above equation, the mean square error is a classification function. The regularization of l_2-norm aims to properly suppress the effect of the classifiers with lower identification abilities on the recognition results in the learning process of classification to ensure that the classification results are more stable. Herein, parameter w selects l_2-norm rather than l_1-norm since regularization lies in that the former plays a less important role than the latter in the regularization of the mean square error function, and the computation of the former is obviously less complex than that of the latter.

$$\|a - Dw\|_2^2 = a\|a - 1 + aw + Dw\|_2^2 = \|[a;1] - [D,a]w\|_2^2 \tag{4}$$

Let $\hat{a} = [a;1]$, $\hat{D} = [D,a]$, then we can get

$$\hat{w} = \arg\min_w (\|\hat{a} - \hat{D}w\|_2^2 + \tau\|w\|_2) \quad s.t.\ w_j > 0, \quad j = 1,\cdots,s \tag{5}$$

So the solution of Eq. (5) is $\hat{w} = (\hat{D}^T\hat{D} + \tau I)^{-1}\hat{D}^T\hat{a}$ [22].

3.6. Multi-classifier Fusion

Given a training sample set of $g_i (i = 1,2,\cdots,k)$ containing k class, the $1_{tst},\ldots,l_{th},\ldots,L_{th}$ ($L = 21$) classifier are formed successively after the set is partitioned. As defined, the optimal weight for the l_{th} classifier is $\hat{w}_l$. Let x_e^l express the l_{th} occluded test sample; $\hat{e}_1^l$ represent the l_{th} reconstructed occlusion; Φ_l represent the l_{th} training sample set and $\hat{\omega}_1^l$ represent the obtained sparse vector of the l_{th} classifier. $r_i^l(x_e^l)$ is the reconstructed residual obtained by reconstructing the l_{th} occluded test sample using the sparse vector of the l_{th} classifier in i_{th} class, that is $r_i^l(x_e^l) = \|x_e^l - \hat{e}_1^l - \Phi_l\,\delta_i(\hat{\omega}_1^l)\|_2$. $r_i^l(x_e^l)$ reflects the possibility of the test sample belonging to i_{th} class. The smaller the value, the more likely that the test sample belongs to the i_{th} class. The valuation of the posterior probability of the l_{th} classifier for defining test samples is defined as follows:

$$\hat{P}_l(g_i | x_e^l) = \frac{1/(r_i^l(x_e^l))^2}{\sum_{n=1}^{k}(1/r_n^l(x_e^l))^2} \tag{6}$$

Valuation of the posterior probability reflects the possibility of the test sample belonging to the i_{th} class. The greater the value is, the more likely that the test sample belongs to the i_{th} class. The following weighted fusion criterion is defined to classify the test sample:

$$identify(x_e) = \arg\max_{1 \le i \le k} \sum_{l=1}^{L} \hat{w}_l\ \hat{P}_l(g_i | x_e^l) \tag{7}$$

4. Experimental Analysis

There are many standard databases available at present. We selected the ORL [31], YaleA [32] and AR [33] databases in controlled environments to test the FR performance of the WMSRC method proposed in this paper. The ORL database contains 100 front images of 10 individuals with different expression and posture. The YaleA database contains 165 images of 15 individuals with different expression and illumination. The AR database is composed of 3288 frontal face images of 126 individuals under various conditions such as expression, illumination, aging, occlusion, etc. For each individual, 26 images are taken in two separate sessions, among which 13 images from one session are shown in Figure 4.

(c) Figure 4. Face Images on AR Database

All images were normalized to 100*100 pixels in the experiment. The baseline SRC [4], BSRC [4], nearest neighbor (NN) [4] and the patch based SRC (MSRC) methods [16] based classifier are used for comparison. The patch size is set as 10×10 (overlap is 5 pixels). The parameter λ used in SRC, BCRC, MSRC and WMSRC are set as 0.001, 0.001, 0.001 and 0.005, respectively. In all the following experiments, the program is run for 20 times on each database and the average results are reported.

4.1. ORL Database

4.1.1. Various Feature Transformations and Classification Methods: The first experiment was carried out to evaluate the correct recognition rate of face images on the ORL database utilizing various feature transformation (dimension reduction) and classification methods. For each individual on the ORL database, 2~3 images were randomly chosen for training and another 5 images for test. The feature dimension was set to be 90 in the experiment. The PCA, LDA, down-sampling and Random method were respectively adopted to reduce the dimension. The size of the block were set to be 25*25 for the BSRC method. The results are shown in Table 1.

Table 1. Correct Recognition Rates (%) on ORL Database

Method	Feature extraction method			
	PCA	LDA	Random	Down-sample
NN	84.1	85.3	**78.4**	71.1
SRC	90.5	91.7	**91.9**	87.3
BSRC	91.7	91.9	**92.3**	91.9
MSRC	93.3	92.9	**94.8**	93.1
WMSRC	93.8	93.7	**95.4**	93.9

When using the same dimensionality reduction method, the recognition rate of the WMSRC method was highest, followed by MSRC, with NN being the lowest, which can be attributed to the dictionary learning based SRC and multi-classifiers weighted optimization. The random sampling matrix method had higher recognition rate compared to the other three feature transformation methods for SRC, BSRC, MSRC and WMSRC method, which was because the transformed matrix of the random method had the lowest correlation coherence degree [2] with the dictionary matrix composed of training samples. Thus, the random method was used in the follow-up experiments for dimensionality reduction.

4.1.2. Various Feature Dimensions and Classification Methods: We tested the impact of different dimensionality reduction modes and classification methods on the ORL database. The training dataset and testing dataset used for the experiment were the same as those used for the first experiment. The random method was used for feature transformation to reduce the dimension to 60, 90, 120, and 250 respectively. The SRC, BSRC, NN, WSRC and WMSRC methods were adopted respectively for classification, where the sizes of the blocks were set at 25*25 for the BSRC method. The results are separately shown in Table 2.

Table 2. Correct Recognition Rates (%) on ORL Database

Method	Feature dimension			
	60	90	120	250
NN	76.8	77.6	78.9	79.0
SRC	86.3	91.8	94.6	95.2
BSRC	86.9	93.2	95.2	96.1
MSRC	87.8	95.6	96.9	97.3
WMSRC	88.1	95.8	97.2	97.5

It can be seen from the results that the correct recognition rate increased with the increase in feature dimension for the four classification methods. The WMSRC achieved the highest recognition rate of 97.3% on the YaleA database, which was due to the combination of SRC based ensemble modeling and voting with weight based multi-classifiers weighted optimization. At the same time, the SRC, BSRC,WSRC, and WMSRC methods achieved a recognition rate above 91.8% on the ORL database and 87.4% on the ORL database when feature dimension was larger than 90. Taking the recognition rate and the cost into consideration, the feature dimension was set at 90 for the subsequent experiment.

4.2. YaleA Database

4.2.1. Various Training Sample Size and Classification Methods: We tested the influence of different training sample size and face recognition on the YaleA database. For each individual on the YaleA database, 2~5 images were randomly chosen for training and another 6~9 images for test. The image dimension was reduced to 90 by the random method. The results on the YaleA database are separately shown in Table 3.

When selecting the same training sample size, the recognition rate of the WMSRC method was highest, followed by MSRC, with NN being the lowest. It can be clearly seen that WMSRC achieves the highest recognition rate on all experiments with the training sample size increasing from 2 to 5. Compared to SRC, BSRC and MSRC, WMSRC leads to much better results, validating the effectiveness of multi-classifiers weighted optimization and dictionary learning based SRC.

Table 3. Correct Recognition Rates (%) on YaleA Database

method	Training Sample size			
	2	3	4	5
NN	83.3	85.1	69.1	77.6
SRC	91.5	91.7	88.3	91.8
BSRC	91.9	92.1	91.8	93.2
MSRC	93.4	93.7	93.9	95.4
WMSRC	94.1	93.9	94.1	95.8

4.2.2. Various Levels of Occlusion and Classification Methods: We tested the influence of different random occlusion sizes on face recognition. We selected the first 2~3 images from the face database to form a training set, and added different occlusion sizes to the last five images to form a testing set. The image dimension was reduced to 90, and occluded face recognition was conducted using SRC, BSRC (block sizes 25*25, 25*50 and 50*50), NN and the proposed method. The results on the YaleA database are separately shown in Table 4.

Table 4. Correct Recognition Rates (%) on YaleA Database

Method	Percentage of occlusion			
	15%	25%	35%	50%
NN	76.8	77.6	78.9	79.0
SRC	83.4	77.8	65.9	41.2
BSRC(25*25)	93.9	89.8	78.2	62.6
BSRC(25*50)	92.0	88.5	73.4	57.7
BSRC(50*50)	85.5	81.3	68.4	45.0
MSRC	95.1	91.7	83.6	67.4
WMSRC	96.8	93.6	85.9	71.3

With the increase in occlusion, the recognition rates of the four algorithms decreased; of them, the recognition rate of the proposed method decreased at the lowest pace, and its recognition rate was highest under the same occlusion. This was because the proposed method was based on dictionary learning and multi-classifier weight optimization, with the highest robustness of occlusion. In addition, under the same occlusion, with the increasing number of blocks, the BSRC recognition rate increased. The recognition rate of the proposed method with the largest number of blocks was highest, far more than the recognition rate of SRC and NN without blocks. This was because BSRC and the proposed method established SRC modes based on partial blocks. The more blocks there were, the greater was the detail. Thus, it was possible to obtain additional local information on images and improve the occlusion recognition rate [17].

4.3. AR Database

The experiment was carried out to evaluate the correct recognition rate of face images on the AR database utilizing various rule fusion and classification methods. In this experiment, we selected a subset with only illumination and expression changes that contains 50 male subjects and 50 female subjects from the AR dataset. For each individual on the subset, 2~3 images from the first session were randomly chosen for training and 1 image from the second session for test. The image dimension was reduced to 90. Sparse representation-based multiple-classifier fusion recognition experiments are implemented based on weighted fusion (WMSRC), sum rule fusion, maximum-value rule fusion, median-value rule fusion and voting rule fusion (MSRC) respectively. The results are separately shown in Table 5.

Table 5. Correct Recognition Rates (%) on AR Database

Fusion Rule	maximum-value	median-value	voting	sum rule
Recognition accuracy	83.3	91.4	90.6	93.2

Experimental results indicate that the WMSRC, which is based on sum rule fusion, has the best recognition performance. This is because the weighted fusion is based on different identification abilities and complies with the subjective perception of human eyes. All the weights are optimal weight vectors through optimization solution. This fusion criterion takes reconstructed residuals into account, thus having better recognition performance than other fusion recognition classifiers.

4.4. Algorithm Efficiency

The experiment was carried out to evaluate the average runtime in classifying a test sample with a gallery matrix from the different database. For each individual on the YaleA and ORL database, we randomly chose 2 images for training, and chose 1 image for test. In this experiment, we selected a subset with only illumination and expression changes that contains 10 male subjects and 10 female subjects from the AR dataset. For each individual on the subset, 2~3 images from the first session were randomly chosen for training and 1 image from the second session for test. The image dimension was reduced to 90. The results are separately shown in Table 6.

It can be seen from the experimental results that, the recognition time of NN algorithm in larger AR database is the shortest, *i.e.*, 0.249s. The recognition time of SRC algorithm is longer, *i.e.*, 0.371s. MSRC divides the face into 21 blocks for weight optimization and voting recognition, thus having a longer recognition time of 0.745s. WMSRC requires not only weight optimization but also fusion, thus, the recognition time thereof is the longest, *i.e.*, 0.792s. This indicates that the sparse representation-based classifier has longer recognition time and lower efficiency compared with the NN classifier. Meanwhile, with the increase of the number of blocks and the increasing complexity of the algorithm, the corresponding running time of CPU of sparse representation-based classifiers becomes longer, and the efficiency becomes lower. The WMSRC algorithm proposed in the present paper has the lowest efficiency. Although the efficiency of this algorithm is low, it can still be processed in real time because that the recognition time in different face libraries is not so long.

Table 6. Average Runtime (s) on Different Databases

Method	YaleA	ORL	AR (The first Session)	AR (The second Session)
SRC	0.020	0.085	0.365	0.371
BSRC	0.034	0.112	0.451	0.458
MSRC	0.056	0.228	0.732	0.745
WMSRC	0.063	0.257	0.792	0.758
NN	0.012	0.058	0.249	0.243

5. Conclusions

In order for a more effective FR when the training samples size is very small, we proposed a multi-classifier weighted optimization and sparse representation based (WMSRC) method for FR. The face image was firstly divided by multi-resolution blocking. The multi-scale SRC classifiers were then established and assigned different weights. According to the multiple voting results of the classifiers, the weights of multi-classifiers were optimized by the least-squares optimization equation with l_2-norm regularization. Finally, the classification results of all blocks were combined according to the weighted fusion criterion. Experiments on the ORL, YaleA and AR database show that the WMSRC method outperforms many existing BSRC methods for FR when the number of training samples per class is very limited. The classifier used by the WMSRC method is the weighted optimization and fusion algorithm, and other classifiers will be used for face recognition in our future research.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (Grant No. 60973064).

References

[1] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face recognition: A literature survey", ACM Computing Surveys, vol. 35, no. 4, **(2003)**.

[2] P. J. Phillips, P. J. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the face recognition grand challenge", Computer vision and pattern recognition, 2005, CVPR 2005. IEEE computer society conference on, IEEE, San Diego, CA, USA, **(2005)** June 20-26.

[3] W. Yu, X. Teng and C. Liu, "Face Recognition Using Discriminate Locality Preserving Projections", Image and Vision Computing, vol. 24, no. 3, **(2006)**.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry and Y. Ma, "Robust face recognition via sparse representation", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 31, no. 2, **(2009)**.

[5] L. Zhang, P. Zhu, Q. Hu and D. Zhang, "A linear subspace learning approach via sparse coding", Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, Barselona, Spain, **(2011)** November 6-11.

[6] X. Tan, S. Chen, Z. H. Zhou and F. Zhang, "Face recognition from a single image per person: A survey, Pattern Recognition", vol. 39, no. 9, **(2006)**.

[7] Y. Su, S. Shan, X. Chen and W. Gao, "Adaptive generic learning for face recognition from a single sample per person", Computer vision and pattern recognition, 2010, CVPR 2010, 2010 IEEE Conference on. IEEE, San Francisco CA, **(2010)** June 13-18.

[8] H. U. Fengsong, "Research for Single Sample Face Recognition Under the Conditions of Pose and Illumination Variant", Hunan University, **(2009).**

[9] R. Kumar, A. Banerjee, B. C. Vemuri and H. Pfister, "Maximizing all margins: Pushing face recognition with kernel plurality", Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, Barselona, Spain, **(2011)** November 6-11.

[10] R. Kumar, A. Banerjee and B. C. Vemuri, "Volterrafaces: Discriminant analysis using volterra kernels", CVPR 2009, IEEE, Miami, Florida, USA, **(2009)** June 20-25.

[11] D. Lin and X. Tang, "Recognize high resolution faces: From macrocosm to microcosm", Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, New York, NY, USA, vol. 2, **(2006)** June 17-22.

[12] Y. Su, S. Shan, X. Chen and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition", Image Processing, IEEE Transactions, vol. 18, no. 8, **(2009)**.

[13] L. Wolf, T. Hassner and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 33, no. 10, **(2011)**.

[14] M. Guillaumin, J. Verbeek and C. Schmid, "Is that you? metric learning approaches for face identification", Computer Vision, 2009 IEEE 12th International Conference on. IEEE, Kyoto, Japan, **(2009)** September 27-October 4.

[15] M. Yang and Z. Lei, "Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary", Proceedings of European Conference on Computer Vision. Crete: ECCV, Crete, Greece, **(2010)** September 5-7.

[16] M. X. Nguyen, Q. M. Le, V. Pham, T. Trung and H. L. Bac, "Multi-scale Sparse Representation for Robust Face Recognition", Proceeding(s) of Conference on Knowledge and Systems Engineering (KSE), Hanoi, Vietnam, **(2011)** Octobor 14-17.

[17] P. Zhu, L. Zhang, Q. Hu and S. C. K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization", Computer Vision–ECCV 2012, Springer Berlin Heidelberg, Florence, Italy, **(2012)** October 7-13.

[18] L. Yong, Y. Jianping, W. Jinhua and L. Wei, "Multibiometric Fusion Based on FAR and FRR", Acta Automatic Sinaca, vol. 37, no. 4, **(2011)**.

[19] M. Aharon, M. Elad and A. Bruckstein, "The K-SVD：an algorithm for designing of overcomplete dictionaries for sparse representation", IEEE Transactions on Signal Processing, vol. 54, no. 11, **(2006)**.

[20] Y. Qing-shan, G. Cheng-an and J. Ming-lu, "Face Recognition Based on Gabor Multi-channel Weighted Optimization and Sparse Representation", Journal of Electronics & Information Technology, vol. 33, no. 7, **(2011)**.

[21] A. Y. Yang, J. Wright, Y. Ma and S. S. Sastry, "Feature Selection in Face Recognition: A Sparse Representation Perspective", Manuscript Submitted to IEEE Trans. PAMI, **(2007)**.

[22] M. Yang, L. Zhang, D. Zhang and W. Shenlong, "Relaxed collaborative representation for pattern classification", Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, Rhode, Island, **(2012)** June 16-21.

[23] M. Yang, L. Zhang, D. Zhang and W. Shenlong, "Relaxed collaborative representation for pattern classification", Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, Rhode, Island, **(2012)** June 16-21.

Authors

Deng Nan, She received the MSc degree (2008) in mathematics from NorthWest University, P.R.CHINA. She is now a PhD candidate at University of Science and Technology Beijing. Her current research interests include pattern recognition, image processing, non-Newtonian mechanical system control, artificial intelligence.

Zheng-guang Xu, He received the PhD degree (2000) in control science and engineering from University of Science and Technology Beijing, Beijing, P.R.CHINA. He is now a professor at University of Science and Technology Beijing. His current research interests include pattern recognition, image processing, non-Newtonian mechanical system control, Petri nets.

Sheng-Qin Bian, She received the MSc degree(2004) in mathematics from Beijing institute of Technology, P.R.CHINA. She is a Ph.D. candidate at University of Science and Technology Beijing. Her current research interests include image modeling and sparse representation, image compression.