International Journal of Innovative Computing, Information and Control
Volume **11**, Number **5**, October **2015**
ICIC International ©2015 ISSN 1349-4198
pp. 1551–1562

ROBUST FACE RECOGNITION FRAMEWORK WITH BLOCK WEIGHTED SPARSE REPRESENTATION BASED CLASSIFICATION

Jun He, Tian Zuo, Bo Sun*, Xuewen Wu, Yongkang Xiao and Xiaoming Zhu

College of Information Science and Technology
Beijing Normal University
No. 19, Xinjiekouwai Street, Haidian District, Beijing 100875, P. R. China
*Corresponding author: tosunbo@bnu.edu.cn

Received January 2015; revised May 2015

Abstract. *The sparse representation based classification method can be divided into two categories: holistic approaches and local feature-based approaches. In spite of the significant success in face recognition, improvements on higher robustness or lower computational complexity are still necessary for its real application. Thus, we first propose a novel Block Weighted Sparse Representation based Classification (BW-SRC) method based on the maximum likelihood model. Then, to ensure the accuracy of BW-SRC, we conduct a pre-alignment process by utilizing the locations of local feature points (in this article, we use SIFT keypoints). Combining the pre-alignment process and BW-SRC, we establish a novel framework for robust face recognition, which is more effective and more robust than the state-of-the-art methods in practical scenarios. Finally, by conducting experiments on AR and Yale databases, the performance of our proposed method and framework is demonstrated and compared with global SRC and blocked SRC. The proposed framework is proven as low-computation, alignment-free and robustness to rotation, illumination and disguise, and more appropriate for practical scenarios.*
Keywords: Sparse Representation based Classification (SRC) method, Block Weighted Sparse Representation based Classification (BW-SRC), Local feature points, Alignment free, Robust face recognition

1. Introduction.

Face Recognition (FR) is one of the hot topics in pattern recognition and computer vision due to its special theoretical research and wide application. Many methods have been proposed for facial recognition, some of which have shown good performance under controlled conditions.

In general, current methods can be divided into two categories according to the features utilized for recognition: holistic approaches and local feature-based approaches [1]. With the help of feature extraction approaches, such as Principle Component Analysis (PCA) [2], Independent Component Analysis (ICA) [3], Locality Preserving Projections (LPP) [4] and Neighborhood Preserving Embedding (NPE) [5], holistic approaches have been widely studied and variant classification methods have been successfully applied to FR, such as Nearest Neighbor (NN) [6], Nearest Subspace (NS) [7], and Support Vector Machine (SVM) [8]. Recently, Sparse Representation based Classification (SRC) [9] method has attracted an increasing amount of attention due to its impressive recognition accuracy and robustness against corruption or occlusion. The principle of SRC is to consider the query face image as a linear combination of training samples with the sparse constraint and then to obtain the recognition result through solving a convex optimization problem for its sparse representation. All of these holistic approaches utilize the information of whole images (often extracted by some dimensionality reduction methods) and have a relatively low computational complexity. Local feature-based approaches address the problem in a different way. They utilize local features for classification. Usually, the extracted local features have some advantages on robustness to different scenarios. For instance, Scale Invariant Feature Transform (SIFT) [10,11] has already been applied in FR tasks [12-14] and has attracted much attention. SIFT descriptors have scale and rotation invariance and show significant robustness in practical FR scenarios. Other local feature extraction approaches, such as PCA-SIFT [15], Speeded up Robust Features (SURF) [16] and Gradient Location-Orientation Histogram (GLOH) [17] have also shown their robust performance against scale, rotation or even affine transformation. Unsurprisingly, SRC has also been introduced to classify face images with these local features, and new methods, such as Multi-Keypoint Descriptor (MKD)-SRC [18], have been proposed.

Unfortunately, no method can perform consistently well in real complex scenarios and each has its own limitations. Holistic approaches, for example, require good alignment to ensure classification accuracy and are weakly robust to contiguous occlusions in real-world scenarios, such as disguise and expression. Furthermore, these approaches do not perform well in the situation of drastic illumination changes. As to some existing local feature-based approaches, such as MKD-SIFT, their huge computational complexity is one of the principal problems. Abundant local features make recognition more accurate, but also increase the computational time. In addition, because some local feature extraction methods (such as SIFT) themselves are sensitive to illumination change, MKD-SRC is usually not an illumination-robust method.

Therefore, it is still essential to find an approach with more robustness against different scenarios in the real world and the relatively lower temporal complexity. We notice that in some conditions with occluded images, using divided blocks of images for classification is better than using whole images. For instance, in the research of [9], for the scenario where face images are occluded by sunglasses or a scarf, the authors partition them into blocks, classify each block independently and combine the classification results in some way (such as majority voting). This method performs better than the approach that uses holistic images for classification. However, they treat occluded and clean blocks equally in the classification process. The research of [19] proposes an idea of weighting the image-blocks with sparsities and residuals of their sparse representations, which helps to discard occluded blocks and preserves clean blocks. However, the blocks still need to be put together to reconstruct a whole image for classification, which reduces flexibility and introduces computational complexity.

Inspired by the above approaches, we propose a method called *Block Weighted Sparse Representation based Classification* (BW-SRC), which classifies each block independently and uses the sparsity and the residual of the sparse representation of each to weight each classification result. The basic idea is: clean blocks should impact the result more than occluded ones. For each block in one image, the probability of the block belonging to each class, rather than the classification result, is computed. In addition, a maximum likelihood approach is used to combine these probabilities to obtain the classification result for the whole image and the weights calculated above are added to the exponents of probabilities. Unlike majority voting, the probability of each block belonging to each class is preserved in the maximum likelihood, making the classification results of whole images more accurate and more robust.

Because the training and query samples in BW-SRC need to be aligned, an appropriate alignment method is needed to be conducted before the subsequent steps. To solve this problem, we utilize the locations of SIFT keypoints on a mean face to align query face images to the same canonical pose [20]. Unlike the approaches relying on the eyeball

[21-25] or other semantic features, this method does not need the appearances of given facial features and can address the images with expressions and occlusions.

The main contributions of this paper are as follows. (1) We propose a novel *Block Weighted Sparse Representation based Classification* (BW-SRC) method, which uses the sparsity and the residual of the sparse representation of each image blocks to determine their importance in classification. In addition, we preserve the probability of each block belonging to each class and use the maximum likelihood model to combine these probabilities to obtain the classification result of the whole image. All of the above helps to make the classification results more accurate and more robust against disguise and illumination variation. (2) We propose a new framework for robust face recognition, which first utilizes the locations of local feature points (such as SIFT keypoints) to align face images in different poses and then recognizes the identity using BW-SRC. Some experiments in our research have indicated that our framework is more effective and more robust than the state-of-the-art methods in practical scenarios, especially the scenarios with non-uniform illumination and disguise.

The rest of this paper is organized as follows. The fundamentals of Sparse Representation (SR) and the common approach of holistic SRC are explained in Section 2. In Section 3, we propose our novel approach, called Block Weighted Sparse Representation based Classification. In Section 4, a preprocessing step to align face images is introduced. Then, we show our whole framework in Section 5. In Section 6, experiments are conducted on AR [26] database and Yale database [27] to demonstrate the performances of our methods. Conclusions and future work are proposed in Section 7.

2. **Sparse Representation based Classification.** In this section, we first briefly introduce the principle of the Sparse Representation based Classification [9] method. In the theory of Sparse Representation (SR), a query sample can be expressed by a linear combination of training samples as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{y}$ is the vectored test sample; $\boldsymbol{A}$ is the dictionary from all training samples of all classes, in which each column is a vectored training sample; $\boldsymbol{x}$ is a sparse coefficient vector in which only a few entries are nonzero. The sparsest solution to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ can be obtained by solving

$$\hat{\boldsymbol{x}}_{\boldsymbol{0}} = \arg\min \|\boldsymbol{x}\|_0 \text{ subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \tag{2}$$

where $\|\cdot\|_0$ denotes the ℓ_0-norm, which is defined as the number of nonzero entries in the vector. Equation (2) is an NP-hard problem, and it has been proven that if the solution $\boldsymbol{x}_{\boldsymbol{0}}$ is sparse enough, the solution of (2) is equal to the solution to the following ℓ_1-minimization problem [28-30]:

$$\hat{\boldsymbol{x}}_{\boldsymbol{1}} = \arg\min \|\boldsymbol{x}\|_1 \text{ subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}. \tag{3}$$

This problem can be solved by greedy pursuit algorithms, such as Orthogonal Matching Pursuit (OMP) [31], or the convex relaxation methods, such as Least Absolute Shrinkage and Selection Operator (LASSO) [32] and Least Angle Regression (LARS) [33]. Once the sparse vector $\boldsymbol{x}$ is recovered, the identity of $\boldsymbol{y}$ can be obtained by the minimal residual:

$$\text{identity}(y) = \arg\min_i \|\boldsymbol{y} - \boldsymbol{A}\delta_i(\hat{\boldsymbol{x}}_{\boldsymbol{1}})\|_2, \tag{4}$$

where $\delta_i(\boldsymbol{x})$ is the characteristic function, which selects only the elements associated with the ith class in $\boldsymbol{x}$.

3. Block Weighted Face Recognition based on Sparse Representation.

3.1. Sparse representation based classification on blocks of images. The holistic images dictionary is denoted as $\boldsymbol{A} \in R^{D\times n}$, where D is the dimension of the vectors of holistic images and n is the number of training images from C subjects. We partitioned each training image into L blocks. The value of L can be 8, as shown in Figure 4 and Figure 6. Thus, the dictionary $\boldsymbol{A}$ is now partitioned as L sub-dictionaries $\{A_k|k = 1, 2, \ldots, L\}$, where $A_k \in R^{d\times n}$, with $d < n$ containing the kth block of each training image. Correspondingly, the query image $\boldsymbol{y}$ is also partitioned into $\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_L}$.

Similar to (1), each block $\boldsymbol{y_k}$ can also be represented as a linear combination of corresponding dictionary $\boldsymbol{A_k}$:

$$\boldsymbol{y_k} = \boldsymbol{A_k}\boldsymbol{x_k}, \quad k = 1, 2, \ldots, L. \tag{5}$$

SR theory is used to obtain the sparse representation $\boldsymbol{x_k}$ of $\boldsymbol{y_k}$, and then, decision rule (4) is used to obtain the classification result of the kth block. Here, we do not simply make a simple decision that the kth block belongs to the mth class but use the representation residual

$$r_k^j = \|\boldsymbol{y_k} - \boldsymbol{A_k}\delta_j(\hat{\boldsymbol{x}}_{\boldsymbol{k}})\|_2 \tag{6}$$

to measure the probability of $\boldsymbol{y_k}$ belonging to the jth class:

$$p_k^j = P(\text{identity}(\boldsymbol{y_k}) = j) = \frac{1/r_k^j}{\sum_{j=1}^{C}\left(1/r_k^j\right)}. \tag{7}$$

3.2. Determination of block weights. As mentioned above, we use the sparsity and the residual [19] of the representation of each block to determine its weight in classification. The function to measure the sparsity of the kth block is as follows:

$$s_k = \frac{\sqrt{C}}{\sqrt{C}-1}\left(\frac{\|\boldsymbol{b_k}\|_2}{\|\boldsymbol{b_k}\|_1} - \frac{1}{\sqrt{C}}\right), \tag{8}$$

where

$$\boldsymbol{b_k} = [b_{k1}, b_{k2}, \ldots, b_{kC}]. \tag{9}$$

b_{kj} is calculated by summing the absolute values of the coefficients belonging to the jth class:

$$b_{kj} = \|\delta_j(\boldsymbol{x_k})\|_1 \tag{10}$$

where $\delta_j(\boldsymbol{x_k})$ denotes the characteristic function, which selects the elements associated with the jth class in $\boldsymbol{x_k}$.

According to (8), when $\boldsymbol{b_k}$ has only one nonzero coefficient, s_k reaches the maximum value 1; as the coefficients of $\boldsymbol{b_k}$ become more uniform, s_k becomes smaller; and when all of the coefficients of $\boldsymbol{b_k}$ are the same nonzero value, s_k reaches the minimum value 0.

The residual of a block can be measured by the ℓ_2-norm:

$$r_k = \|\boldsymbol{A_k}\boldsymbol{x_k} - \boldsymbol{y_k}\|_2. \tag{11}$$

The sparsities and the residuals of blocks are effective measurements to discriminate the clean blocks from the occluded ones. Specifically, in previous research, it was found that clean blocks tend to have a large value of sparsity and small value of residual. This is because they can be relatively accurately represented by only a few training samples from the same class. On the contrary, blocks that are completely or partly occluded have small sparsities or large residuals. Therefore, we define the weight function in reference to [19]:

$$w_k = w_k^s w_k^r \tag{12}$$

where

$$w_k^s = \begin{cases} 0, & s_k \leq s_0 \\ (s_k - s_0)/(s_1 - s_0), & s_0 < s_k < s_1 \\ 1, & s_k \geq s_1 \end{cases} \tag{13}$$

$$w_k^r = \begin{cases} 1, & r_k \leq r_1 \\ (r_0 - r_k)/(r_0 - r_1), & r_1 < r_k < r_0 \\ 0, & r_k \geq r_0 \end{cases} \tag{14}$$

s_0, s_1 and r_0, r_1 are soft thresholds, which not only can discriminate occluded and clean blocks effectively but also can have robustness for face images from different databases or applications.

3.3. **Combining the classification results according to block weights.** We have obtained the classification probabilities of blocks and their weights by Equation (7) and Equation (12). Finally, we use the maximum likelihood approach to combine these classification probabilities and add the weights to their exponents. Thus, the identity of a query image $\boldsymbol{y}$ is given by

$$\text{identity}(y) = \arg\max_{i=1,\dots,C} \log\left(\prod_{k=1}^{L} (p_k^i)^{w_k}\right) = \arg\max_{i=1,\dots,C} \sum_{k=1}^{L} w_k \log(p_k^i). \tag{15}$$

With the block weights, the maximum likelihood model can make clean blocks have more effect on the final result while occluded blocks have less effect, which effectively improves the robustness against occlusion. Furthermore, this model can preserve the probability of each block belonging to each class, making the combined classification results more accurate and more robust. The framework of the proposed approach is shown in Algorithm 1.

Algorithm 1: Block Weighted Sparse Representation based Classification

Input: A set of training images partitioned into L blocks $\boldsymbol{A_1}, \boldsymbol{A_2}, \dots, \boldsymbol{A_L}$ with unit ℓ_2-norm column, a query image partitioned into 8 blocks $\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_L}$ with unit ℓ_2-norm column.

1) Solve the block ℓ_1-minimization problem with LARS.
2) Compute the probability of the kth block belonging to the jth class with Equation (7).
3) Compute the sparsity s_k and the residual r_k of each block with Equations (8) and (11).
4) Compute the weight w_k of each block with Equations (12), (13) and (14). If all of the block weights are 0, set $w_k = 1$ for $k = \arg\min_k s_k/r_k$.

Output: identity($\boldsymbol{y}$) $= \arg\max_{i=1,\dots,C} \log\left(\prod_{k=1}^{L} (p_k^i)^{w_k}\right)$

4. **Alignment with Feature Points Positions.** The proposed Block Weighted Sparse Representation based Classification (BW-SRC) method performs well on the condition that face images are well aligned. However, in most real-world scenarios, face images are usually unaligned. Therefore, we conduct a pre-alignment process. We conduct three steps to align the training and the query samples: building a SIFT keypoint set with the common face template, determining the correspondences of keypoints and face alignment by 2-D spatial transformation.

4.1. **Building a SIFT keypoint set with the common face template.** Given a dictionary of pre-aligned training samples, our task is to align the query samples by pose and scale to the training ones. First, we build a common face template based on the dictionary. The mean face $\boldsymbol{m}$ is computed, which captures the common information among different subjects. SIFT keypoints are detected in $\boldsymbol{m}$ to reveal the locations of the common and stable features in the training samples, as shown in Figure 1. $P = \{p_i\}$, $i = 1, \ldots, t$, denotes the set of these keypoints. However, as $\boldsymbol{m}$ is the mean face and is smoothed among subjects, the descriptors extracted in $\boldsymbol{m}$ provide little information and cannot be used in the next alignment steps. Therefore, we detect SIFT keypoints and compute SIFT descriptors directly from the individual training images. The keypoint coordinates detected in $\boldsymbol{m}$ are used to select these keypoints. In each training image, we detect keypoints in the small areas near each location of p_i. If there is more than one keypoint in the neighborhood of p_i, the nearest one is selected. In this way, around each keypoint in p_i, we obtain a series of keypoints N_i. Note that the locations and descriptors of N_i are used in the alignment process rather than the location or descriptor of p_i. Moreover, to make the number of keypoints in N_i less dependent to the number of training images, hierarchical clustering is used to group the descriptors of N_i into n clusters. The cluster centers are selected as the keypoints whose descriptors have the minimum cosine distance to all of the other descriptors in the same cluster. We save the coordinates and descriptors of all of these $t \times n$ keypoints. Hence, the SIFT keypoint set T is built from the training images, which will be used in the next step.

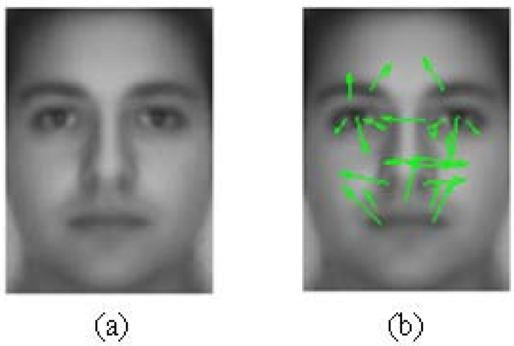
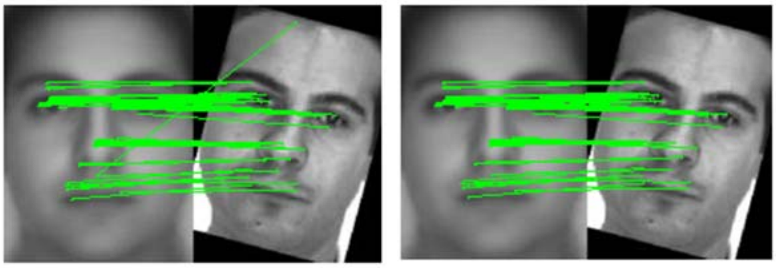


FIGURE 1. (a) Mean face and (b) SIFT keypoints in it

4.2. **Determining the correspondences of keypoints.** We now have the keypoint set T extracted from the training face images. Our probe images, which are supposed to be the output of a face detector, are usually non-aligned and should be aligned to the same canonical pose as the training samples. We denote one probe image as I and the SIFT keypoint set S is extracted from I. The best match of a probe keypoint in S is found by searching its nearest neighbor in the keypoint set T. The nearest neighbor is defined as the keypoint whose descriptor has the maximum similarity to the probe keypoint. After the keypoint match pairs are found between S and T, Hough Transform is used to check their geometric consistencies and reject the false match pairs. There are three parameters for each SIFT keypoint: location, scale and orientation. In our experiments, the orientation bin size used in the Hough transform is 30°, the scale bin size is 2 and the location bin size is 0.25 times the size of the images. The bin, which is passed by the most match pairs, is determined to be the key bin and the match pairs passing this bin have geometric consistencies and are considered to be true keypoint match pairs. Using the methods



(a) (b)

FIGURE 2. The examples of (a) original match pairs and (b) the examples of corrected match pairs

above, we obtain the keypoint pairs in S and T. The examples of original match pairs and corrected match pairs are shown in Figure 2.

4.3. **Face alignment by 2-D spatial transformation.** Suppose (x'_p, y'_p) is the coordinate of a keypoint in the probe image and (x_p, y_p) is its match point in the training keypoint set T. The 2D transformation from (x_p, y_p) to (x'_p, y'_p) can be described as

$$\begin{bmatrix} x'_p \\ y'_p \end{bmatrix} = \begin{bmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{bmatrix} \begin{bmatrix} x_p \\ y_p \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{16}$$

Thus, when we obtain more than two match pairs, the transformation parameters $[s,\ \theta,\ t_x,\ t_y]$ can be computed as follows:

$$\begin{bmatrix} s\cos\theta \\ s\sin\theta \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} x'_{p1} & -y'_{p1} & 1 & 0 \\ y'_{p1} & x'_{p1} & 0 & 1 \\ x'_{p2} & -y'_{p2} & 1 & 0 \\ y'_{p2} & x'_{p2} & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_{p1} \\ y_{p1} \\ x_{p2} \\ y_{p2} \end{bmatrix}. \tag{17}$$

Once the transformation parameters are obtained, we can transform the probe image I according to the 2-D transformation.

5. **Robust Face Recognition Framework.** Utilizing the alignment method and BW-SRC approach, we can build a robust face recognition framework. First, we conduct the alignment process to align the query images to the training samples. Then, BW-SRC is used to obtain the robust classification results with the block dictionaries and image blocks obtained by partitioning the aligned query images. The flowchart of our proposed framework is shown in Figure 3.

6. **Experiments.** In this section, we apply the proposed BW-SRC method and the alignment method on the publicly available AR and Yale databases to evaluate their performances. Further, the performance of the whole framework is demonstrated and compared with the state-of-the-art methods.

6.1. **Face images with disguise and illumination variation.** First, we construct the training and probe sets with images of the two databases, respectively. In the AR database, images of 100 subjects are utilized, among which there are 50 males and 50 females. For each subject, 8 images with varying facial expressions and uniform illumination are applied as training samples. The face images are resized as 120×160px and then partitioned into 8 blocks as shown in Figure 4(a). Each block has $40 \times 60 = 2400$ pixels. Random Projection is used to reduce the dimension of each block to 120. To test our approach robustness against disguise and illumination variation, 200 images (2 images per

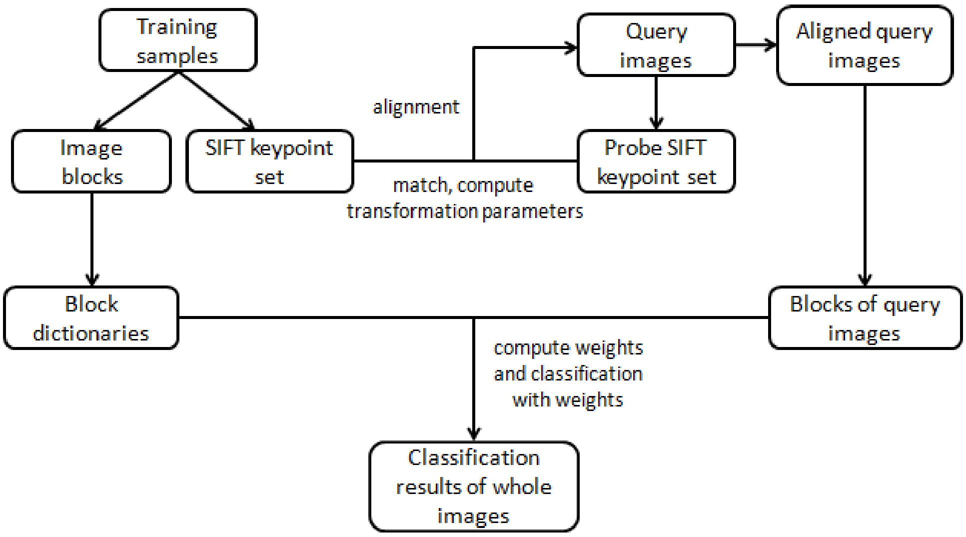


FIGURE 3. The flowchart of the robust face recognition framework

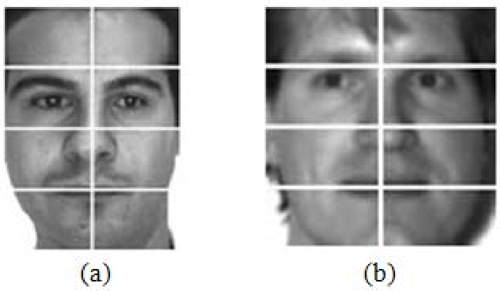


FIGURE 4. Partiton scheme of training samples in (a) AR database and (b) Yale database

subject) with sunglasses, 200 images (2 images per subject) with a scarf and 400 images (4 images per subject) with non-uniform illumination compose the probe set, which are partitioned and dimensionality reduced in the same way as that of the training images, as shown in Figure 6(a). Some examples of the probe images are shown in Figure 5(a). Similarly, in the Yale database, images of 14 subjects are utilized, which were cropped, aligned and resized as 100×100px. 112 face images (8 images per subject) under varying facial expressions with uniform illumination compose the training set, each of which is partitioned into 8 blocks, as shown in Figure 4(b). Each block has $25 \times 50 = 1250$ pixels. Random Projection is used to reduce the dimension of each block to 100. The probe set is composed of 14 images (1 image per subject) with glasses and 28 images (2 images per subject) with non-uniform illumination. Some examples of probe images are shown in Figure 5(b), and their partition scheme is shown in Figure 6(b).

Next, BW-SRC is performed on the image sets. The probability of each query image block belonging to each class is calculated with the SRC method and (7). The convex relaxation method Least Angle Regression (LARS) [33] is used to find the sparse representation for the image blocks. Then, the block weights and the maximum likelihood model are used to determine the classification results of whole images based on the block classification results. The recognition rates of the query images in the AR database and Yale database are shown in Table 1 and Table 2, respectively, compared with the conventional approaches, global SRC and block SRC [9]. For global SRC, which utilizes holistic images

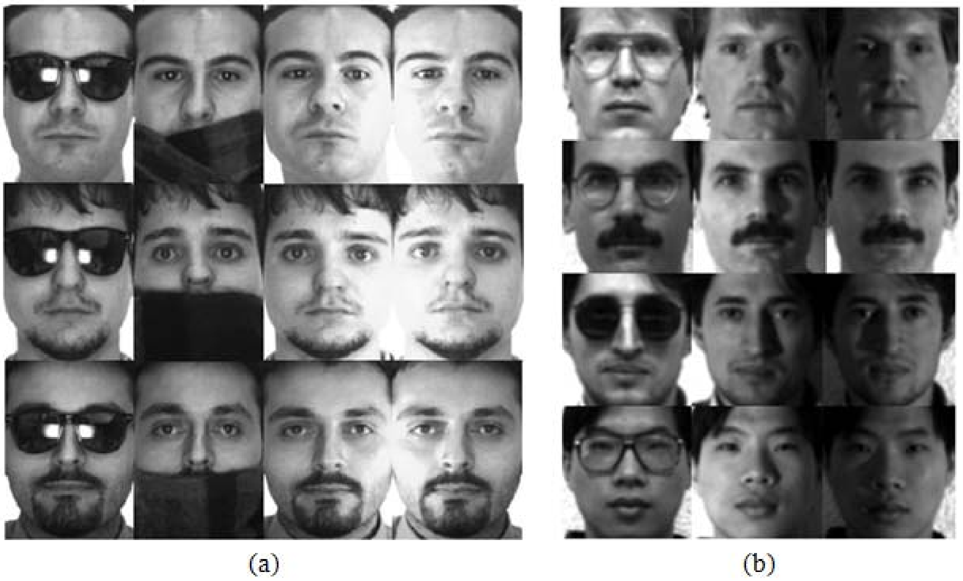


FIGURE 5. The examples of probe images in (a) AR database and (b) Yale database

FIGURE 6. Partiton scheme of probe images in (a) AR database and (b) Yale database

TABLE 1. Recognition rates on AR database

Scenarios	**Sunglasses**	**Scarf**	**Non-Uniform Illumination**
global SRC	66.50%	16.00%	69.50%
block SRC	92.50%	92.50%	98.00%
BW-SRC	**98.50%**	**97.50%**	**99.25%**

TABLE 2. Recognition rates on Yale database

Scenarios	**Disguise**	**Non-Uniform Illumination**
global SRC	100%	60.71%
block SRC	92.86%	75.00%
BW-SRC	100%	**92.86%**

for recognition, we reduce the dimensions of the images to 240 for the AR database and 100 for the Yale database.

As is shown in Table 1 and Table 2, the recognition rates of BW-SRC are significantly higher than global SRC and block SRC in the scenarios with disguise (mainly sunglasses or a scarf) and non-uniform illumination.

6.2. **Performance of the alignment approach and the effectiveness of our proposed framework.** To test the performance of the alignment approach and the effectiveness of our proposed framework, we artificially produce non-alignment with the query images in the AR database. 200 images with a clean face and 200 images with a scarf are transformed randomly with one of the two groups of transformation, $[s = 0.9, \theta = \pi/12, t_x = -27, t_y = 4]$ and $[s = 0.9, \theta = -\pi/12, t_x = 11, t_y = -25]$. Then, the alignment approach is conducted to align them to the training samples. The examples of query images with artificial transformation and their alignment results are shown in Figure 7. Then, BW-SRC is used to recognize the identities of these aligned images, and the final recognition rates are shown in Table 3 compared with the state-of-the-art method, MKD-SRC. The computation time is calculated by a computer with an Intel Core2 CPU E7200, 2.00 GB RAM and Windows 7 (64-bit).

FIGURE 7. The examples of query images with artificial transformation and their alignment results. (a) Images of clean face. (b) Images with sunglasse and scarf.

TABLE 3. Recognition rates of proposed framework

Scenarios	Clean Face	Scarf	Computation Time (average per sample)
proposed framework	98.5%	84.25%	**0.77s**
MKD-SRC	100%	96.5%	9.37s

As is shown in Table 3, the clean faces' recognition rate of our framework is comparable to the MKD-SRC method in [18], while the recognition rate of the images with a scarf is relatively lower. It is especially notable that our framework requires much less computation time than MKD-SRC (the time consumption of MKD-SRC is a dozen of times that of our proposed approach).

7. **Conclusions and Future Work.** In this paper, we optimized the block-SRC method and proposed an optimized approach and a practical framework for face recognition. Using the sparsity and the residual of the sparse representation, we have obtained the importance of image blocks for recognition. This factor is elaborately combined into a new decision rule, a maximum likelihood model, to classify a query sample. Thus, we propose a novel *Block Weighted Sparse Representation based Classification* (BW-SRC) method. Considering practical scenarios, we address the problem of alignment by utilizing the locations of local feature points (such as SIFT keypoints). Combining this preprocessing

step and the proposed BW-SRC, we have obtained a practical framework for robust face recognition. To verify the effectiveness of the proposed method and framework, we performed experiments on publicly available face-databases, including the AR and Yale data sets. The experimental results show that the BW-SRC based framework achieves demonstrable improvement in classification performance compared with the global or block SRC method. The experimental results and performance comparisons confirm that the proposed BW-SRC based framework can robustly address the image classification problem in disguise and non-uniform illumination scenarios. Additionally, in practical scenarios, the recognition rates of our framework are comparable to the MKD-SRC method, while our framework requires much less computation time.

In our future research, we intend to improve the performance of alignment in the preprocessing step by utilizing Harris corner points to find local feature points, and the invariance to scale can be achieved by utilizing other local feature descriptors, such as PCA-SIFT, SURF or GLOH, instead of SIFT descriptor. Additionally, some statistic moments, such as the Hu Moment invariants, may also be used as local feature descriptors in the framework. These researches on feature points and feature descriptors may help to improve the robustness of our framework in the situation of non-uniform illumination.

Acknowledgement. This work was supported by the Fundamental Research Funds for the Central Universities (2014KJJCA15), the Specialized Research Fund for the Doctoral Program of Higher Education (20120003120038), and the Fundamental Research Funds for the Central Universities (2012LYB48), the State Key Laboratory of Acoustics, Chinese Academy of Sciences (SKLA201304), the National Natural Science Foundation of China (61431004), and the Fundamental Research Funds for the Central Universities (2013NT55).

REFERENCES

[1] C. Geng and X. Jiang, Face recognition based on the multi-scale local image structures, *Pattern Recognition*, vol.44, no.10, pp.2565-2575, 2011.
[2] S. Wold and K. Esbensen, Geladi P.: Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, vol.2, no.1, pp.37-52, 1987.
[3] P. Comon, Independent component analysis, *Higher-Order Statistics*, pp.29-38, 1992.
[4] X. Niyogi, Locality preserving projections, *Neural Information Processing Systems*, vol.16, no.153, 2004.
[5] X. He, D. Cai, S. Yan et al., Neighborhood preserving embedding, *The 10th IEEE International Conference on Computer Vision ICCV*, vol.2, pp.1208-1213, 2005.
[6] T. Cover and P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Information Theory*, vol.13, no.1, pp.21-27, 1967.
[7] J. Ho, M. H. Yang, J. Lim et al., Clustering appearances of objects under varying illumination conditions, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp.11-18, 2003.
[8] C. Cortes and V. Vapnik, Support vector machine, *Machine Learning*, vol.20, no.3, pp.273-297, 1995.
[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31, no.2, pp.210-227, 2009.
[10] D. G. Lowe, Object recognition from local scale-invariant features, *Proc. of the 7th IEEE International Conference on Computer Vision*, vol.2, pp.1150-1157, 1999.
[11] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol.60, no.2, pp.91-110, 2004.
[12] M. Bicego, A. Lagorio, E. Grosso et al., On the use of SIFT features for face authentication, *Conference on Computer Vision and Pattern Recognition Workshop CVPRW'06*, p.35, 2006.
[13] J. Luo, Y. Ma, E. Takikawa et al., Person-specific SIFT features for face recognition, *IEEE International Conference on Acoustics Speech and Signal Processing ICASSP*, vol.2, pp.593-596, 2007.

[14] C. Geng and X. Jiang, Face recognition using sift features, *The 16th IEEE International Conference on Image Processing (ICIP)*, pp.3313-3316, 2009.
[15] Y. Ke and R. Sukthankar, PCA-SIFT: A more distinctive reresentation for local image descriptor, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, vol.2, pp.506-513, 2004.
[16] H. Bay, T. Tuytelaars and L. Van Gool, Surf: Speeded up robust features, *Computer Vision – ECCV*, pp.404-417, 2006.
[17] K. Mikolajczyk and C. Schmid, A performance evaluation of local descriptors, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.10, pp.1615-1630, 2005.
[18] S. Liao and A. K. Jain, Partial face recognition: An alignment free approach, *International Joint Conference on Biometrics (IJCB)*, pp.1-8, 2011.
[19] J. Lai and X. Jiang, Modular weighted global sparse representation for robust face recognition, *IEEE Signal Processing Letters*, vol.19, no.9, pp.571-574, 2012.
[20] C. Geng and X. Jiang, Fully automatic face recognition framework based on local and global features, *Machine Vision and Applications*, vol.24, no.3, pp.537-549, 2013.
[21] Y. W. Chen and K. Kubo, A robust eye detection and tracking technique using gabor filters, *The 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, vol.1, pp.109-112, 2007.
[22] T. D'Orazio, M. Leo, G. Cicirelli et al., An algorithm for real time eye detection in face images, *Proc. of the 17th International Conference on Pattern Recognition*, vol.3, pp.278-281, 2004.
[23] G. Li, X. Cai, X. Li et al., An efficient face normalization algorithm based on eyes detection, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.3843-3848, 2006.
[24] D. T. Lin and C. M. Yang, Real-time eye detection using face-circle fitting and dark-pixel filtering, *IEEE International Conference on Multimedia and Expo ICME*, vol.2, pp.1167-1170, 2004.
[25] C. W. Park, J. M. Kwak, H. Park et al., An effective method for eye detection based on texture information, *International Conference on Convergence Information Technology*, pp.586-589, 2007.
[26] A. M. Martinez, The AR face database, *CVC Technical Report 24*, 1998.
[27] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19, no.17, 1997.
[28] D. L. Donoho, For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution, *Communications on Pure and Applied Mathematics*, vol.59, no.6, pp.797-829, 2006.
[29] E. J. Candes, J. K. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, vol.59, no.8, pp.1207-1223, 2006.
[30] E. J. Candes and T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies, *IEEE Trans. Information Theory*, vol.52, no.12, pp.5406-5425, 2006.
[31] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, *Proc. of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, pp.40-44, 1993.
[32] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B (Methodological)*, pp.267-288, 1996.
[33] B. Efron, T. Hastie, I. Johnstone et al., Least angle regression, *The Annals of Statistics*, vol.32, no.2, pp.407-499, 2004.