Modular Weighted Global Sparse Representation for Robust Face Recognition

Jian Lai and Xudong Jiang, *Senior Member, IEEE*

***Abstract*—This work proposes a novel framework of robust face recognition based on the sparse representation. Image is first divided into modules and each module is processed separately to determine its reliability. A reconstructed image from the modules weighted by their reliability is formed for the robust recognition. We propose to use the modular sparsity and residual jointly to determine the modular reliability. The proposed framework advances both the modular and global sparse representation approaches, especially in dealing with disguise, large illumination variations and expression changes. Compared with the related state-of-the-art methods, experimental results on benchmark face databases verify the advancement of the proposed method.**

***Index Terms*—Face recognition, contiguous occlusion, modular representation, sparse representation.**

I. Introduction

FACE recognition has attracted a lot of researchers due to its wide application. Many methods have been proposed to solve this problem, such as Principle Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [2], Independent Component Analysis (ICA) [3], Eigenfeature Regularization and Extraction (ERE) [4] and Support Vector Machines (SVM) [5]. Recently, Sparse Representation Coding (SRC) [6], which considers the query face image as a linear combination of all training samples with sparse constrain, is receiving more and more attention. SRC harnesses l_1-norm to approach l_0-norm, which achieves impressive recognition accuracy. Similar to the sparse constrain, Linear Regression Classification (LRC) [7] casts the task as a linear combination of samples from just a single subject.

Although these methods perform well under controlled conditions, they cannot handle contiguously occluded face images in the real-world scenario, such as disguise shown in Fig. 1(a). Therefore, modular approach [8] was applied in both SRC and LRC. They partition the face image into a number of modules. Each module is processed separately. The final decision is made by fusing the classification result of each module. The modular SRC [6] uses majority voting for classification. The disadvantage of majority voting is that it treats occluded and clean modules equally. In modular LRC [7], authors classify query face image by labeling it to the subject with the minimum class representation error in all modules. However, the limited discriminating power of a single module tends to cause misclassification.

Manuscript received March 22, 2012; revised May 18, 2012; accepted June 25, 2012. Date of publication July 05, 2012; date of current version July 13, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Deniz Erdogmus.

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Link, Singapore 639798 (e-mail: jlai1@ntu.edu.sg; exdjiang@ntu.edu.sg).

Digital Object Identifier 10.1109/LSP.2012.2207112

Fig. 1. The first row shows normal images. The second row shows image variations caused by disguise, expression and large illumination changes.

The commonality of the above methods is that they essentially treat each module independently. Hence, they lose the correlation information between modules which is very important for classification.

These observations motivate us to explore a more robust approach for face recognition. Firstly, classification should only rely on the clean modules. Secondly, the correlation information between the clean modules should be utilized in the recognition process. Based on these two criterions, we propose a modular weighted global sparse representation (WGSR) method. It is composed of two stages. First, image is divided into modules and each module is processed separately. The modular sparsity and residual are jointly employed to determine the reliability of each module. Then, the modular reliability is used to weight the module for the reconstruction of a global feature vector. Classification is performed on the reconstructed global feature vector.

Besides disguise, the proposed method can also handle large illumination variations and expression changes. While caused by different sources, these variations share some common properties. First, they all have some pixels largely different from those of normal images. Moreover, the distinct pixels are contiguous that form one or a few local areas. As shown in Fig. 1, regions in the green boxes are very different from normal images in the first row. The distinct area provides little discriminating or even misleading information for classification. In this work, we try to solve these three typical problems under a common framework that tackles the contiguous occlusion.

II. Proposed Approach

Given a set of n training images from c different subjects, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$, where $\mathbf{a}_i$ is the feature vector representing the ith image. Each training image is partitioned into eight modules as shown in Fig. 2, which is the same face image partition choice as [6], [7]. Correspondingly, $\mathbf{A}$ is partitioned as $\mathbf{A} = [\mathbf{A}_1^T, \mathbf{A}_2^T, \ldots, \mathbf{A}_8^T]^T$ and $\mathbf{A}_k \in R^{d\times n}$ with $d < n$. Obviously, the lth column of $\mathbf{A}_k$ is the feature vector for the kth module of the lth image. Similarly, a query image $\mathbf{y}$ is partitioned into $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_8$.

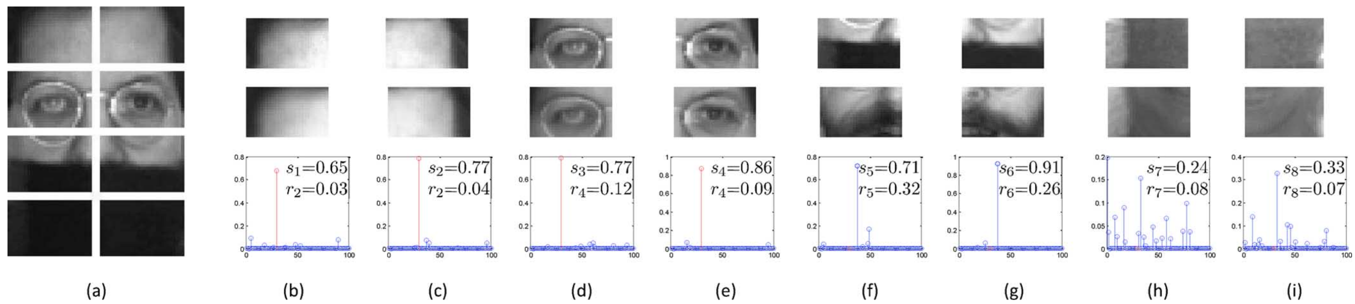

Fig. 2. (a) Query face image with scarf. The first row of (b)–(i) shows the eight modules of query image. The second row of (b)–(i) shows the reconstructed image with l_1-norm minimization. The third row of (b)–(i) shows the estimated modular class coefficients, the numeric values of modular sparsity and residual. The red coefficients correspond to correct training subject.

A. Determination of the Modular Weights

For each module $\mathbf{y}_k$, we treat it as a linear combination of the dictionary $\mathbf{A}_k$ by

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k. \tag{1}$$

As this is an underdetermined liner system $(d < n)$, there are infinitely many solutions. To get a unique and stable solution, a constrain is needed. The most popular constrain is l_2-norm minimization for its simplicity. However, it generates a dense solution that is not informative for recognizing the query image. To achieve a sparse solution, in [6], l_1-norm minimization is employed.

Apply l_1-norm minimization, we get the following optimization problems

$$\mathbf{x}_k = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{A}_k\mathbf{x} - \mathbf{y}_k\|_2^2 + \lambda\|\mathbf{x}\|_1 \right\} \quad \text{for } 1 \leq k \leq 8. \tag{2}$$

The sparsity should be only measured between classes [9]. Thus, we propose to add up the absolute values of coefficients belonging to the same class to form the modular class coefficient b_{kj} for the kth module of the jth class

$$b_{kj} = \|\delta_j(\mathbf{x}_k)\|_1, \tag{3}$$

where $\delta_j(\mathbf{x}_k)$ is the characteristic function which selects the elements associated with the jth class in $\mathbf{x}_k$.

Concatenate all c modular class coefficients to generate the modular class vector $\mathbf{b}_k$ as

$$\mathbf{b}_k = [b_{k1}, b_{k2}, \ldots, b_{kc}]. \tag{4}$$

We propose the following function for measuring the modular sparsity:

$$s_k = \frac{\sqrt{c}}{\sqrt{c}-1}\left(\frac{\|\mathbf{b}_k\|_2}{\|\mathbf{b}_k\|_1} - \frac{1}{\sqrt{c}}\right). \tag{5}$$

When $\mathbf{b}_k$ has only a single nonzero coefficient, s_k reaches the maximum value 1. When all coefficients of $\mathbf{b}_k$ are the same nonzero value, s_k reaches the minimum value 0.

The modular residual can be measured simply by the l_2-norm as

$$r_k = \|\mathbf{A}_k\mathbf{x}_k - \mathbf{y}_k\|_2. \tag{6}$$

After generating the above measurements, we explore the relation of the occlusion with the modular sparsity and residual. Clean query modules can be accurately represented by only the training samples from the same class. Therefore, they produce large value of the sparsity and small value of the residual. This is verified by the numeric values of $s_1 - s_4$ and $r_1 - r_4$ as shown in Fig. 2(b)–(e). The situations of the occluded modules are more complicated. An occluded module could be far away from every subspace spanned by training samples of each class but near or even within the subspace spanned by all training samples of all classes. This results in a small value of the sparsity s_k but also a small value of residual. This is verified by the numeric values of s_7, s_8 and r_7, r_8 as shown in Fig. 2(h) and (i). While this kind of occluded modules can have even smaller residuals than the clean ones, they can be differentiated by the small values of the sparsity. However, the sparsity alone cannot differentiate all kinds of occluded modules. If an occluded module is significantly nearer the subspace of one class than the others, the optimization (2) will produce, though a large residual, a high value of the sparsity. This is verified by the numeric values of s_5, s_6 and r_5, r_6 as shown in Fig. 2(f) and (g). Therefore, to differentiate occluded modules from the clean ones, both the modular sparsity and residual should be employed.

A simple way to tackle the occlusion problem is to exclude the occluded modules from the feature vector for the classification. This can be done by multiplying all modules with a weighting function w_k as

$$w_k = \begin{cases} 0, & \text{if } (s_k \leq s_t) \vee (r_k \geq r_t) \\ 1, & \text{if } (s_k > s_t) \wedge (r_k < r_t). \end{cases} \tag{7}$$

Obviously, we can compute the weighting function simply by $w_k = w_k^s w_k^r$, where w_k^s and w_k^r are two step functions, respectively switching from 0 to 1 at s_t and from 1 to 0 at r_t.

However, it is very difficult to find the optimal hard thresholds s_t and r_t, which could be different from different databases or applications. In addition, it is not difficult to understand that the final classification performance will be very sensitive to the two hard thresholds. Therefore, we relax each hard threshold to two easily found safe thresholds $s1, s2$ and $r1, r2$ so that $(s_k \geq s2) \wedge (r_k \leq r1)$ guarantees a clean module and $(s_k \leq s1) \vee (r_k \geq r2)$ ensures a significantly occluded module. We simply weight modules between these two extreme cases linearly to s_k and r_k. Thus, the weighting function is proposed as

$$w_k = w_k^s w_k^r, \tag{8}$$

where

$$w_k^s = \begin{cases} 0, & s_k \leq s1 \\ (s_k - s1)/(s2 - s1), & s1 < s_k < s2 \\ 1, & s_k \geq s2 \end{cases} \tag{9}$$

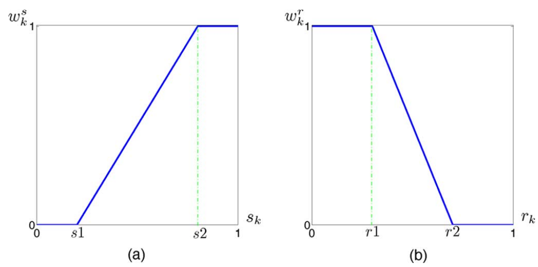

Fig. 3. (a) Weighting function of modular sparsity. (b) Weighting function of modular residual.

$$w_k^r = \begin{cases} 1, & r_k \le r1 \\ (r2 - r_k)/(r2 - r1), & r1 < r_k < r2 \\ 0, & r_k \ge r2 \end{cases}. \quad (10)$$

Fig. 3 shows the two weighting values w_k^s and w_k^r as the functions of the sparsity s_k and residual r_k, respectively.

B. Weighted Global Sparse Representation

The modular representation based methods make a final decision based on the classification result of each module. The loss of the correlation information between modules may cause misclassification. We propose to combine the weighted modules to reconstruct a global feature vector.

Each modular training matrix $\mathbf{A}_k$ and modular query vector $\mathbf{y}_k$ are weighted by corresponding weighting value to generate a modular weighted training matrix and a modular weighted query vector:

$$\mathbf{A}_k^w = w_k \mathbf{A}_k, \quad (11)$$
$$\mathbf{y}_k^w = w_k \mathbf{y}_k. \quad (12)$$

Concatenate all modular weighted training matrices to produce a weighted global training matrix as

$$\mathbf{A}^r = \left[\mathbf{A}_1^{wT}, \mathbf{A}_2^{wT}, \dots, \mathbf{A}_8^{wT} \right]^T. \quad (13)$$

Concatenate all modular weighted query vectors to produce a weighted global query vector as

$$\mathbf{y}^r = \left[\mathbf{y}_1^{wT}, \mathbf{y}_2^{wT}, \dots, \mathbf{y}_8^{wT} \right]^T. \quad (14)$$

We consider that $\mathbf{y}^r$ is a linear combination of $\mathbf{A}^r$ as

$$\mathbf{y}^r = \mathbf{A}^r \mathbf{x}^r. \quad (15)$$

To capture the subject label information from the weighted global matrix and weighted global query vector, we solve the following global l_1-norm minimization problem

$$\mathbf{x}^r = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{A}^r \mathbf{x} - \mathbf{y}^r\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}. \quad (16)$$

It is not difficult to see that $\mathbf{x}^r$ is affected by the modular weighting value w_k. If the heavy weighted module is not well represented, the error will be large. To minimize the global reconstruction error, $\mathbf{x}^r$ tends to reconstruct the modules with larger weights more accurately. Therefore, the reliable modules are heavily weighted. The unreliable modules are weighted by small value down to zero to reduce their impacts on the classification result.

C. Classification Procedure

Given a weighted global query vector $\mathbf{y}^r$, a weighted global training matrix $\mathbf{A}^r$ and the representation coefficient vector $\mathbf{x}^r$, we allocate the query image to the subject i that has the global minimum residual

$$i = \arg\min_j \|\mathbf{A}^r \delta_j(\mathbf{x}^r) - \mathbf{y}^r\|_2. \quad (17)$$

The following Algorithm 1 summarizes the proposed recognition framework.

Algorithm 1: Modular Weighted Global Sparse Representation (WGSR)

Input: A set of training images partitioned into eight modules $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_8$ with unit l_2-norm column. A query image partitioned into 8 modules $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_8$ with unit l_2-norm column.

1) Solve the modular l_1-norm minimization problems as (2).
2) Compute the modular sparsity s_k and residual r_k by (5) and (6).
3) Compute the modular weights w_k by (8), (9), and (10). If all modular weighting values equal 0, set w_k as 1 for $k = \arg\min_k\{(r_k)/(s_k)\}$.
4) Reconstruct the weighted global training matrix $\mathbf{A}^r$ and weighted global query vector $\mathbf{y}^r$ by (11), (12), (13), and (14).
5) Normalize each reconstructed image to unit l_2-norm column. Solve the global l_1-norm minimization problem as (16).

Output: $\mathrm{identify}(\mathbf{y}) = \arg\min_j \|\mathbf{A}^r \delta_j(\mathbf{x}^r) - \mathbf{y}^r\|_2$.

Compared with the global LRC and SRC approaches, the proposed WGSR framework attenuates the problems of the query images with corrupted, occluded or largely varied modules that may mislead the representation and classification. Compared with the modular based LRC and SRC approaches, the proposed WGSR framework solves the problems of the limited discriminating information of each individual module by seeking the global optimization instead of multiple local optimizations in the modular based LRC and SRC approaches.

III. Experiments

We evaluate the performance of the proposed method on the same databases as in [6]: Extended Yale B and AR database. These two face databases focus on frontal faces with illumination variations, disguises and expression changes. In the experiments, we compare the proposed method with the related methods: global SRC [6], modular SRC [6], global LRC [7] and modular LRC [7].

For the proposed method, suitable parameters are studied by cross-validation on the unused 26 subjects of AR database (see Section III-B). $\lambda = 0.05$, $r1 = 0.05$, $r2 = 0.2$, $s1 = 0.2$, and $s2 = 0.8$ are chosen and fixed in all experiments.

TABLE I
FACE RECOGNITION RATE ON THE EXTENDED YALE B DATABASE

Image size	10×12	16×20	20×24
global SRC	95.1%	97.1%	97.7%
modular SRC	67.3%	97.0%	98.4%
global LRC	**95.2**%	96.3%	96.6%
modular LRC	6.2%	73.4%	87.5%
WGSR	93.7%	**98.9**%	**99**%

TABLE II
FACE RECOGNITION RATE ON THE AR DATABASE

Scenarios	Sunglass	Scarf	Expression
global SRC	87.0%	59.5%	84.3%
modular SRC	97.5%	93.5%	83.8%
global LRC	65.5%	12.5%	85.2%
modular LRC	93.5%	73.5%	82.3%
WGSR	**100**%	**97.5**%	**92.4**%

A. Extended Yale B Database

The cropped Extended Yale B database consists of 2414 frontal-face images of 38 subjects with size 168×192, captured under 64 different lighting conditions. Same as in [6], we randomly select half of the images of each subject for training and the rest are used for testing. To test the effect of the image size on the recognition accuracy, the images are down-sampled to the sizes of $10 \times 12 = 120$, $16 \times 20 = 320$, and $20 \times 24 = 480$. Table I shows the average face recognition rates over ten runs for different image sizes across various methods.

For the small image size of 10×12, the performances of the modular approaches are much worse than the global methods due to the insufficient features (dimensionality of 15) for each module. Nonetheless, the recognition rate of the WGSR approach, though utilizing modular information, is much better than the modular methods and close to the global methods. The increase of the image size to 16×20 significantly improves the performances of the modular approaches. However, the advantages of the modular approaches do not yet fully compensate their disadvantages. The proposed WGSR approach outperforms both the modular and global methods. For the large image size of 20×24, the performances of the modular approaches are further improved and the modular SRC outperforms its global counterpart. Still, the proposed WGSR framework achieves the best recognition accuracy.

B. AR Database

The AR database consists of over 4000 frontal-face images from 126 subjects. For each subject, 26 pictures were taken in two separate sessions. In the experiment, same as in [6], 50 male subjects and 50 female subjects are selected.

In the first set of experiment, we use 799 unoccluded images (about 8 per subject) under varying facial expression as training samples with image size of $60 \times 80 = 4800$. The 4800 dimensional feature vector is directly processed without down-sampling as in [6]. We consider two separate disguise test sets of 200 images. The first disguise set contains images of the subjects wearing sunglasses, which occludes roughly 20% of the image. The second disguise set contains images of the subjects wearing a scarf, which occludes roughly 40% of image.

In the second set of experiment, we test the face recognition on AR database with lower resolution. 200 images of neutral expression (two per subject) are chosen for training. 600 images with facial expressions (smile, anger and scream) are selected for testing. All images are down-sampled to the feature vector of the dimensionality of $20 \times 28 = 560$.

Table II lists the recognition performances. For the two disguise scenarios, as the query samples are severely occluded, the recognition rates of the global methods are much lower than those of the modular approaches. In the expression changes scenario, however, the modular approaches are inferior to the global ones. One reason could be that the expression changes causes small amount of variation but they spread over most modules. As a result, the advantages of the modular approaches do not fully compensate their problems. In all three scenarios, the proposed WGSR framework significantly outperforms both the modular and global approaches.

IV. CONCLUSION AND FUTURE WORK

In this work, a novel framework of robust face recognition is proposed that utilizes the modular approach to reconstruct a weighted global feature vector for sparse representation. Classification is performed on the reconstructed global representation of the face image. This framework attenuates problems of both the modular and global representation methods. We evaluate the proposed method on several conditions, including variation of illumination, expression changes and two different disguises. The experimental results clearly and consistently show that the proposed framework is much more robust than both the modular and global representation methods.

In case the occlusion affects all pre-partitioned modules, repartition of the image is a possible solution. One interesting future work is to develop a mechanism that optimizes the number of modules and the way to partition the image based on the specific problem and images in hand.

REFERENCES

[1] X. D. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 31, no. 5, pp. 931–937, May 2009.

[2] X. D. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 16–26, Mar. 2011.

[3] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, Apr. 1994.

[4] X. D. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, Mar. 2008.

[5] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Int. Conf. Computer Vision*, 2001, vol. 2, pp. 688–694.

[6] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[7] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.

[8] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 84–91.

[9] X. T. Yuan and S. C. Yan, "Visual classification with multi-task joint sparse representation," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2010, pp. 3493–3500.