Manifold Regularized Local Sparse Representation for Face Recognition

Lingfeng Wang, Huaiyu Wu, and Chunhong Pan

***Abstract*— Sparse representation-(or sparse coding)-based classification has been successfully applied to face recognition. However, it can become problematic in the presence of illumination variations or occlusions. In this paper, we propose a Manifold Regularized Local Sparse Representation (MRLSR) model to address such difficulties. The key idea behind the MRLSR method is that all coding vectors in sparse representation should be group sparse, which means holding the two properties of both individual sparsity and local similarity. As a consequence, the face recognition rate can be considerably improved. The MRLSR model is optimized by the modified homotopy algorithm, which keeps stable under different choices of the weighting parameter. Extensive experiments are performed on various face databases, which contain illumination variations and occlusions. We show that the proposed method outperforms the state-of-the-art approaches and provides the highest recognition rate.**

***Index Terms*— Face recognition, manifold regularization, sparse representation.**

I. Introduction

FACE recognition [1], [2] is a critical component in many computer vision applications, such as access control, video surveillance, and public security. In the real world, face images are often corrupted by many unknown factors, namely illumination, occlusion and expression, so face recognition is still a challenging task. In the past decades, a number of researchers have been attracted to tackle these difficulties [3]–[15] and have proposed many effective methods. Among them, sparse representation-based method is one of the promising branches. In this paper, we first briefly review the existing face recognition methods, and then we introduce the sparse representation-based face recognition methods.

A. Brief Introduction of Face Recognition

Until now, many face recognition methods have been proposed. From face representation view, these approaches can be mainly classified into two groups, i.e., the subspace-based holistic methods and the feature-based local methods. It is worth noting that both holistic and local methods can adopt the sparse representation technique.

The effectiveness of subspace-based face recognition approaches depends on the utilization of subspace learning methods, which include the well-known principal component analysis (PCA) [16] and Fisher Linear Discriminate Analysis (FLDA) [17], as well as the manifold learning-based locality preserving projection (LPP) [18], local discriminant embedding [19], and graph embedding [20]. For example, the classical FLDA seeks an optimal linear transformation by maximizing the ratio of between-class scatter and within-class scatter. The manifold learning-based methods consider that the high-dimensional data are often embedded on a low-dimensional manifold. For example, the LPP in [18] learns the subspace by preserving the geometric graph of the original high-dimensional data. Yan *et al.* [20] utilize graph embedding theory to reinterpret the subspace learning methods. Recently, the kernel techniques, such as kernel PCA and kernel LDA [21], are proposed to tackle the nonlinearity in face feature space.

The holistic methods often fail to identify the faces with local variations, such as illumination and expression. Fortunately, the feature-based local face recognition approaches can tackle these difficulties better. Two types of features and their variants, i.e., local binary pattern (LBP) [22] and Gabor [23]–[25], are often used in practice. For example, LBP is a texture primitive statistic, which is obtained by the sign of the difference between center pixel and its neighborhood. It is invariant to linear transformation and is, thereby, robust to illumination variation to some extent. Gabor filters could effectively extract the local structure features at multiple directions and scales. The Gabor feature is also robust to illumination and expression variations.

B. Sparse Representation-Based Face Recognition

Sparse representation has been widely used in both computer vision and signal processing applications. It was first introduced into face recognition by Wright *et al.* [6]. In this method, a testing image is first sparsely represented over all training images, and then the classification result is obtained by finding the class that leads to the minimal representation error. Improved performances have been reported as compared with some classical methods. However, the trivial template represented by identity occlusion dictionary results in high computational cost. In [25], the computational cost is reduced by introducing Gabor features [23], which makes the occlu-

Manuscript received February 12, 2014; revised April 23, 2014 and June 9, 2014; accepted June 30, 2014. Date of publication July 8, 2014; date of current version April 2, 2015. This work was supported in part by the National Natural Science Foundation of China under Grants 61175025 and 61272049 and in part by the Beijing Natural Science Foundation under Grant 4132075. This paper was recommended by Associate Editor S. Yan.

The authors are with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: lfwang@nlpr.ia.ac.cn; hywu@nlpr.ia.ac.cn; chpan@nlpr.ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2014.2335851

sion dictionary compressible. In [9] and [12] robust sparse representation is proposed to avoid the utilization of identity occlusion dictionary, so that the speed of sparse representation can be accelerated. The robust sparse representation method can provide higher recognition rate, especially when the face image is occluded by disguise. In [26], multiple features are combined under a multitask sparse representation framework by introducing a matrix $\ell_{2,1}$-norm [27]. The key idea behind the matrix $\ell_{2,1}$-norm is to ensure the coding vectors are similar to each other. Motivated by the matrix $\ell_{2,1}$-norm regularization, Yang *et al.* [13] proposed a relaxed regularization. The testing and training images are first divided into multiple blocks, and then each block of a testing image is sparsely represented by the corresponding block of the training images. The relaxed regularization is utilized to ensure all the coding vectors are globally similar to each other.

C. Our Method

In this paper, we propose a manifold regularization to ensure that the neighboring coding vectors are similar to each other if they have strong correlation. That is, manifold regularization is proposed to transfer the local correlation of face patches to the local similarity of coding vectors. By incorporating manifold regularization into local (or block-based) sparse representation, we propose a new Manifold Regularized Local Sparse Representation (MRLSR) model. The MRLSR model is optimized by the modified homotopy algorithm, where the updating direction and the updating step are improved by further considering manifold regularization. The weighting parameter of sparse regularization starts from the infinity, and then decreases to a specified value.

Specifically, the advantages of the MRLSR model are highlighted as follows.

1) By introducing manifold regularization, all the coding vectors obtained by the MRLSR are group sparse [28], [29], which means holding the two properties of both individual sparsity and local similarity. As reported in previous work, group sparsity of coding vectors can improve face recognition rate. Extensive experiments also indicate that the proposed MRLSR method provides better adaptability to face recognition difficulties, such as illumination variations and occlusions, compared with the state-of-the-art approaches.
2) Due to the homotopy-like optimization, the proposed MRLSR model keeps stable under different choices of the weighting parameter of sparse regularization. Face recognition results on the AR and PIE database also verify this (please refer to Section V-E).
3) The proposed MRLSR model not only provides high recognition rate, but also is stable under different block size selections. The comparative experiments on various face databases verify these two points.

The rest of this paper is organized as follows. A brief review of the sparse representation-based face recognition is introduced in Section II. In Section III, we propose the manifold regularization and the MRLSR model. The homotopy-like optimization of our model is described in Section IV. Some experimental results are provided in Section V. The concluding remarks are given in Section VI.

II. Brief Review of Sparse Representation

Let $\mathbf{X} = [\mathbf{X}^1; \mathbf{X}^2; \ldots; \mathbf{X}^{\mathrm{C}}]$ be the set of training images, in which $\mathbf{X}^i$ represents images of class i, and C is the number of classes. In sparse representation-based face recognition, the testing image $\mathbf{y}$ is sparsely coded on $\mathbf{X}$ via ℓ_1 minimization, given by

$$\hat{\alpha} = \arg\min_{\alpha} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\alpha\|_2^2 + \gamma_{\text{global}}\|\alpha\|_1 \tag{1}$$

where γ_{global} is a weighting constant, which gives a tradeoff between the reconstruction error and the sparsity of the coding vector α. The objective of recognition is then to find the smallest reconstruction error of all classes, given by

$$\textsc{Identity}(\mathbf{y}) = \arg\min_{i} \|\mathbf{y} - \mathbf{X}^i\hat{\alpha}^i\|_2^2 \tag{2}$$

where $\hat{\alpha} = [\hat{\alpha}^1; \hat{\alpha}^2; \ldots; \hat{\alpha}^{\mathrm{C}}]$, and $\hat{\alpha}^i$ is the coding vector associated with the ith class.

As shown in (1), the optimal solution $\hat{\alpha}$ is sparse. Wright *et al.* [6] denote that the sparsest representation is naturally discriminative: among all subsets of base vectors, it selects the subset which most compactly expresses the input signal and rejects all other possible but less compact representations. In other words, the sparsity could improve discrimination. In object classification problem (face recognition can be regarded as an object classification problem), the discriminative nature is very effective. Hence, the core idea behind sparse representation-based classification is that a query sample should be classified to the class which could faithfully represent it using fewer numbers of samples.

III. Manifold Regularized Local Sparse Representation (MRLSR)

A. Local Sparse Representation

Real face images are often corrupted by partial occlusion and local illumination variation. In such cases, it is fallacious to utilize holistic template to describe face image. Therefore, we present the local sparse representation to solve the above two difficulties. The testing and training images are first divided into $\mathrm{B} = m \times n$ blocks, respectively (with half overlap in practice). Then, for each block of testing image $\mathbf{y}_b$, its coding vector is obtained by optimizing the following sparse representation problem, given by

$$\hat{\alpha}_b = \arg\min_{\alpha_b} \frac{1}{2}\|\mathbf{y}_b - \mathbf{X}_b\alpha_b\|_2^2 + \gamma_b\|\alpha_b\|_1 \tag{3}$$

where $b \in \{1, 2, \ldots, \mathrm{B}\}$ is the block index. Fig. 1 shows the details about block sparse representation. In (3), the blocking strategy effectively introduces spatial information into local sparse representation. Unfortunately, each coding vector α_b is obtained individually. That is, the relationship of all coding vectors $\{\alpha_b\}_{b=1}^{\mathrm{B}}$ is ignored. In practice, it is reasonable to assume that the neighboring blocks should share similar coding vectors. The manifold regularization proposed in Section III-B considers local similarity of the coding vectors.

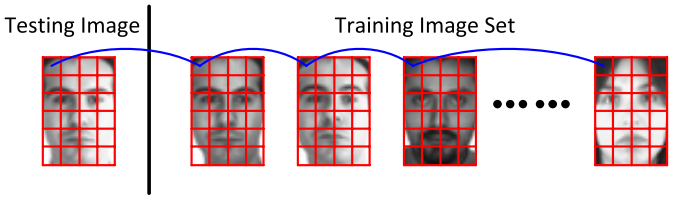

Fig. 1. Local (or block-based) sparse representation. The testing and training images are first divided into $m \times n$ blocks, respectively (see the red grids). Then, each block of the testing image is sparsely represented by the corresponding blocks of the training images (see the blue curves).

B. Manifold Regularization

Assuming that the input image is divided into $\mathrm{B} = m \times n$ blocks, we denote the block index set by

$$\mathcal{B} = \{1, 2, \ldots, \mathrm{B}\}. \tag{4}$$

For the bth block, we consider a 3×3 neighboring block centered on this block as its neighborhood. The neighboring coding vectors of α_b within this local window are

$$\{\alpha_{b_1}, \alpha_{b_2}, \ldots, \alpha_{b_8}\} \tag{5}$$

where $b_1, b_2, \ldots, b_8$ are the coordinates of the blocks arranged from the top-left to the bottom-right. Note that the block itself is not its neighborhood. Similar to Laplacian Eigen maps [30], the weighted difference is utilized to interpret local similarity between α_b and its neighboring coding vectors

$$E(b) = \sum_{i=1}^{8} w_{bb_i} \|\alpha_b - \alpha_{b_i}\|_2^2 \tag{6}$$

where w_{bb_i} is the weight (or similarity) between two blocks. In this paper, the weight w_{bb_i} is obtained by calculating the Canonical Correlation Analysis (CCA) of two corresponding blocks, i.e., $\mathbf{X}_b$ and $\mathbf{X}_{b_i}$.[1]

Denote $\mathbf{W} \in \mathbb{R}^{\mathrm{B} \times \mathrm{B}}$ is a similarity matrix constructed by all weights, that is the bcth element of $\mathbf{W}$ is defined as

$$\mathbf{W}_{bc} = \begin{cases} w_{bc}, & \text{if } c \text{ is the neighbor of } b \\ 0, & \text{otherwise} \end{cases}$$

note that if c is the neighbor of b, c belongs to the set $\{b_i\}_{i=1}^{8}$. Denote $\mathbf{D}$ is a diagonal matrix, satisfying that $\mathbf{D}_{bb} = \sum_c \mathbf{W}_{bc}$. Matrix $\mathbf{L}$ is a Laplacian matrix defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{7}$$

Laplacian matrix $\mathbf{L}$ is a symmetric matrix since the neighborhood system and weight calculation are both symmetric.

[1]In our model, we need a similarity measurement to calculate the similarity between two different set facial parts. Note that each set contains N facial parts, where N is the number of training samples. Most of traditional similarity measurements derive from the Sum of Square Difference (SSD). However, using SSD requires that two samples should be exactly aligned. It is not satisfied here because we need to measure the similarity between different facial parts. Fortunately, the CCA can just satisfy this problem. CCA can be seen as the problem of finding basis vectors for two sets of variables such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Based on the above analysis, we choose CCA but not the other similarity measurements. Refer to [31] for more information about CCA.

With the above definitions, the total difference, which is defined by the summation of all blocks, is

$$\begin{aligned} E(\Lambda) &= \frac{1}{2} \sum_{b=1}^{\mathrm{B}} E(b) \\ &= \frac{1}{2} \sum_{b=1}^{\mathrm{B}} \sum_{i=1}^{8} w_{bb_i} \|\alpha_b - \alpha_{b_i}\|_2^2 = \mathrm{tr}(\Lambda \mathbf{L} \Lambda^{\mathsf{T}}) \end{aligned} \tag{8}$$

in which matrix Λ contains all representation coding vectors, namely, $\Lambda = [\alpha_1; \alpha_2; \ldots; \alpha_{\mathrm{B}}] \in \mathbb{R}^{d \times \mathrm{B}}$ where d is the number of pixels in each block, and tr(.) is the matrix trace operation.

C. MRLSR Model

The manifold regularization on all coding vectors is proposed in (8). Combined with (3), the MRLSR model is defined as

$$F(\Lambda) = \sum_{b=1}^{\mathrm{B}} \left(\frac{1}{2} \|\mathbf{y}_b - \mathbf{X}_b \alpha_b\|_2^2 + \gamma_b \|\alpha_b\|_1 \right) + \frac{\lambda}{2} \mathrm{tr}(\Lambda \mathbf{L} \Lambda^{\mathsf{T}}) \tag{9}$$

where λ is a weighting constant. As shown in (9), the first term is composed of B L_1 regularized SR models, which means that the first term of $F(\Lambda)$ term is convex. Moreover, the Laplacian matrix $\mathbf{L}$ is semidefinite, which indicates that the second term of $F(\Lambda)$ is also convex. Accordingly, we can obtain that the objective function of $F(\Lambda)$ is convex. The core idea behind our MRLSR model is to ensure all coding vectors hold the following two properties: individual sparsity and local similarity. Based on the two properties, the coding vectors are group sparse.

IV. Optimization

The homotopy-like algorithm is used to optimize the proposed MRLSR model. Let $F_\gamma(\Lambda)$ be the objective function with a given sparse regularization coding vector γ, where $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_{\mathrm{B}}]$. Since the objective function is convex, the minimization of $F_\gamma(\Lambda)$ can be achieved when $\Lambda_\gamma = [\alpha_{1,\gamma_1}; \alpha_{2,\gamma_2}; \ldots; \alpha_{\mathrm{B},\gamma_{\mathrm{B}}}]$. Based on convex analysis, the zero vector $\mathbf{0}$ should be an element of the subdifferential at α_{b,γ_b}, given by

$$\begin{aligned} \mathbf{0} = \partial_\Lambda F_\gamma(\alpha_{b,\gamma_b}) = &-\left(\mathbf{X}_b^{\mathsf{T}}(\mathbf{y}_b - \mathbf{X}_b \alpha_{b,\gamma_b}) - \lambda \Lambda_\gamma \mathbf{L}_b\right) \\ &+ \gamma_b \partial \|\alpha_{b,\gamma_b}\|_1 \quad \forall b \in \mathrm{B} \end{aligned} \tag{10}$$

where $\mathbf{L}_b$ is the bth column of the Laplacian matrix $\mathbf{L}$, and $\mathbf{u} = \partial \|\alpha_{b,\gamma_b}\|_1$ is the subgradient, satisfying that

$$\begin{cases} \mathbf{u}(j) = \mathrm{sgn}(\alpha_{b,\gamma_b}(j)), & \alpha_{b,\gamma_b}(j) \neq 0 \\ \mathbf{u}(j) \in [-1, 1], & \alpha_{b,\gamma_b}(j) = 0 \end{cases} \tag{11}$$

where sgn(.) is the sign function and (j) means the jth item of a vector. Let

$$\mathbf{c}_b = \mathbf{X}_b^{\mathsf{T}}(\mathbf{y}_b - \mathbf{X}_b \alpha_{b,\gamma_b}) - \lambda \Lambda_\gamma \mathbf{L}_b$$

be the residual vector, and $\mathcal{I}_{\gamma_b} = \{j : \alpha_{b,\gamma_b}(j) \neq 0\}$ be the active set of the coding vector α_{b,γ_b}. Equation (10) is equivalent to the following two conditions:

$$\textit{Condition 1:} \quad \mathbf{c}_b(j) = \gamma_b \mathrm{sgn}(\alpha_{b,\gamma_b}(j)) \; j \in \mathcal{I}_{\gamma_b} \tag{12}$$

$$\textit{Condition 2:} \quad |\mathbf{c}_b(j)| \leq \gamma_b, \text{ otherwise} \tag{13}$$

similar to the homotopy [32], our algorithm starts from the zero initial solution, i.e., the initial coding vectors $\{\alpha_b^0\}_{b=1}^{\mathrm{B}}$, and then updates them iteratively $\{\alpha_b^t\}_{b=1}^{\mathrm{B}},\ t = 1, 2, \ldots$.

A. Updating Direction

For each coding vector $\alpha_{b,\gamma_b},\ b \in \mathcal{B}$, the corresponding updating direction $\mathbf{e}_b(\mathcal{I}_{\gamma_b})$ in the active set $\mathcal{I}_{\gamma_b}$ is obtained by solving the following equation:

$$\mathbf{X}_b(\mathcal{I}_{\gamma_b})^{\mathsf{T}}\mathbf{X}_b(\mathcal{I}_{\gamma_b})\mathbf{e}_b(\mathcal{I}_{\gamma_b}) + \lambda\mathbf{E}(\mathcal{I}_{\gamma_b})\mathbf{L}_b = \mathrm{sgn}(\alpha_b(\mathcal{I}_{\gamma_b})) \quad (14)$$

where $\mathbf{E}(\mathcal{I}_{\gamma_b}) = [\mathbf{e}_1(\mathcal{I}_{\gamma_b}); \mathbf{e}_2(\mathcal{I}_{\gamma_b}); \ldots; \mathbf{e}_{\mathrm{B}}(\mathcal{I}_{\gamma_b})]$. The updating direction ensures that (12) should be still held after updating. In (14), the updating directions of all coding vectors are coupled together. To solve it effectively, the Landweber iteration [33] is utilized. That is, $\mathbf{e}_b(\mathcal{I}_{\gamma_b})$ is obtained by iteratively solving the following function:

$$\begin{aligned}\mathbf{e}_b^t(\mathcal{I}_{\gamma_b}) = {} & (\mathbf{X}_b(\mathcal{I}_{\gamma_b})^{\mathsf{T}}\mathbf{X}_b(\mathcal{I}_{\gamma_b}) + \lambda\mathbf{D}_{bb}\mathbf{Id})^{-1} \\ & \left(\mathrm{sgn}(\alpha_b(\mathcal{I}_{\gamma_b})) + \lambda\mathbf{E}^{t-1}(\mathcal{I}_{\gamma_b})\mathbf{W}_b\right) \end{aligned} \quad (15)$$

where $\mathbf{E}^{t-1}$ is the previous updating direction matrix, $\mathbf{e}_b^t$ is the current bth updating direction vector, and $\mathbf{Id}$ is the identical matrix. In each iteration, the updating direction vectors in the active set $\mathcal{I}_{\gamma_b}$ are set to $\mathbf{e}_b(\mathcal{I}_{\gamma_b})$, while those not in the active set $\mathcal{I}_{\gamma_b}$ are set to zero.

B. Updating Step

The updating step $\Delta\gamma$ is determined by calculating a minimum step that breaks the following two conditions. First, a nonactive element of $\mathbf{c}_b(j)$ would increase beyond γ, given by

$$\gamma_b^+ = \min_{j\in\mathcal{I}_{\gamma_b}^c}\left\{\frac{\gamma_b - \mathbf{c}_b(j)}{1 - \mathbf{X}(j)^{\mathsf{T}}\mathbf{X}(\mathcal{I}_{\gamma_b})\mathbf{e}_b(\mathcal{I}_{\gamma_b}) - \lambda\mathbf{E}(\mathcal{I}_{\gamma_b})\mathbf{L}_b}, \frac{\gamma_b + \mathbf{c}_b(j)}{1 + \mathbf{X}(j)^{\mathsf{T}}\mathbf{X}(\mathcal{I}_{\gamma_b})\mathbf{e}_b(\mathcal{I}_{\gamma_b}) + \lambda\mathbf{E}(\mathcal{I}_{\gamma_b})\mathbf{L}_b}\right\} \quad (16)$$

where $\mathcal{I}_{\gamma_b}^c = \mathcal{B} - \mathcal{I}_{\gamma_b}$. Second, for an element in the active set, its coding value equals to zero, given by

$$\gamma_b^- = \min_{j\in\mathcal{I}_{\gamma_b}}\left\{\frac{-\alpha_{b,\gamma_b}(j)}{\mathbf{e}_b(j)}\right\}. \quad (17)$$

The updating step $\Delta\gamma_b$ is obtained by

$$\Delta\gamma_b = \min_b\{\gamma_b^+, \gamma_b^-\}. \quad (18)$$

Equation (18) restricts that only one condition can be broken when updating the coding vectors. When $\gamma_b^+ < \gamma_b^-$, we add an element i^+, the minimal index in (16), to the active set. Otherwise, we delete an element i^-, the minimal index in (17), from the active set.

C. Summarization of the Algorithm

Following the above procedure, the updating directions $\{\mathbf{d}_b\}_{b=1}^{\mathrm{B}}$ and their corresponding steps $\{\Delta\gamma_b\}_{b=1}^{\mathrm{B}}$ are obtained. Thus, the regularization parameters $\{\gamma_b\}_{b=1}^{\mathrm{B}}$ are updated by

$$\gamma_b^t = \gamma_b^{t-1} - \Delta\gamma_b \quad \forall b \in \mathcal{B} \quad (19)$$

Algorithm 1 MRLSR

Data: Training images $\{\mathbf{X}_b\}_{b=1}^{\mathrm{B}}$, testing image $\{\mathbf{y}_b\}_{b=1}^{\mathrm{B}}$, the weighting constant λ, and the threshold $\epsilon = 10^{-3}$.
Result: The coding vectors $\{\alpha_b\}_{b=1}^{\mathrm{B}}$.

```
Constructing the Laplacian matrix L based on (7);
Initializing the coding vectors {α_b^0 = 0}_{b=1}^B;
Initializing the γ_b with max{X_b^T y_b};
for t = 0 to maxIterOuter do
    for i = 0 to maxIterInner do
        Updating the directions {d_b}_{b=1}^B by Eqn. (15);
    end
    for b = 1 to B do
        Calculating γ_b^+ and γ_b^- by (16) and (17);
        Computing updating step Δγ by (18);
        if γ_b^+ < γ_b^- then
            Adding i^+ to I_{γ_b} : I_{γ_b} = I_{γ_b} + i^+;
        else
            Removing i^- from I_{γ_b} : I_{γ_b} = I_{γ_b} - i^-;
        end
        Updating regularization parameter by (19);
        Updating the coding vector by (20);
        if min{{γ_b}_{b=1}^B} < ε then
            Break;
        end
    end
end
```

and the coding vectors $\{\alpha_b\}_{b=1}^{\mathrm{B}}$ are updated by

$$\alpha_b^t = \alpha_b^{t-1} + \Delta\gamma_b\mathbf{e}_b \quad \forall b \in \mathcal{B} \quad (20)$$

where t is the iteration number. Details of MRLSR are summarized in Algorithm 1.

As shown in Algorithm 1, the complexity of the outer loop is $O(\mathrm{TB})$, where T is the iteration number and B is the block number. The computational complexity in the inner loop is $O(\mathrm{TN}^3)$, where N is the number of training samples. Therefore, the computational complexity of our algorithm is $O(\mathrm{TB}^2\mathrm{N}^3)$.

The parameter ϵ is utilized to control the weighting parameter of the sparse regularization. Without being specified, it is set to 10^{-3}. It is worth noting that in our algorithm, the weighting parameter of the sparse regularization starts from the infinity, and then decreases to the specified value ϵ. As a result, our algorithm keeps stable under different choices of the weighting parameter ϵ (Section V-E).

After all coding vectors $\{\hat{\alpha}_b\}_{b=1}^{\mathrm{B}}$ are obtained by Algorithm 1; the recognition is determined by finding the smallest reconstruction error of all classes, given by

$$\textsc{Identity}(\mathbf{y}) = \arg\min_i \sum_{b=1}^{\mathrm{B}} \|\mathbf{y}_b - \mathbf{X}_b^i\hat{\alpha}_b^i\| \quad (21)$$

where $\hat{\alpha}_b = [\hat{\alpha}_b^1; \hat{\alpha}_b^2; \cdots; \hat{\alpha}_b^{\mathrm{C}}]$, and $\hat{\alpha}_b^i$ is the coding vector associating the bth block with ith class.

V. Experimental Results

Extensive experiments on three public face databases, i.e., AR, Extended Yale B, and CMU-PIE, are performed to

Fig. 2. Image samples from the AR database. The first row illustrates 7 training images with only illumination and expression changes, while the second and third rows show images with the sunglass and scarf occlusions, respectively.

TABLE I
RECOGNITION RATE COMPARISONS WITH THE RCR WITH DIFFERENT BLOCK SIZES ON THE AR DATABASE WITHOUT OCCLUSION

Algorithm \ Block Size	12*16	12*8	8*16	8*8
Local SR	95.43%	97.29%	96.00%	95.72%
RCR	96.71%	96.00%	96.14%	93.14%
MRLSR	**98.86%**	**98.86%**	**99.00%**	**99.00%**

evaluate the effectiveness of the proposed MRLSR. In our algorithm, the major parameter λ is experimentally set to 10^{-2}.

We compare our method with the state-of-the-art sparse representation-based face recognition approaches, including sparse representation without using occlusion dictionary [6] (SR), local sparse representation without regularization (Local SR),[2] relaxed collaborative representation [13] (RCR), robust sparse representation [9] (RSR), and tone aware sparse representation [14] (TaSR).[3] For SR, TaSR, and Local SR, the homotopy algorithm is utilized to optimize the ℓ_1 problem. For RCR and RSR, the source codes are available at the URL www4.comp.polyu.edu.hk/~slzhang/code.htm. We utilize the default well-tuned parameters for the RCR and RSR face recognition algorithms.

A. Results on the AR Database Without Occlusion

In the first experiment, we test our MRLSR method on the AR database without occlusion [34]. Similar to [6], a subset of images with only illumination and expression variations is utilized (see the first row of Fig. 2). All images are cropped and resized to 48 $\times$ 32. For each person, the seven images from Session 1 are selected as the training images, while other seven images from Session 2 are chosen as the testing images.

The comparative results with the RCR method are shown in Table I. From this table, we obtain three observations. First, our results are better than those of the RCR and the Local SR. Second, our results are stable under different selections of the block size. Third, when the block size is smaller, the superiority of our result is more obvious. Especially, when the block size is 8×8, the recognition rate of our MRLSR is 5.86% higher than that of the RCR.

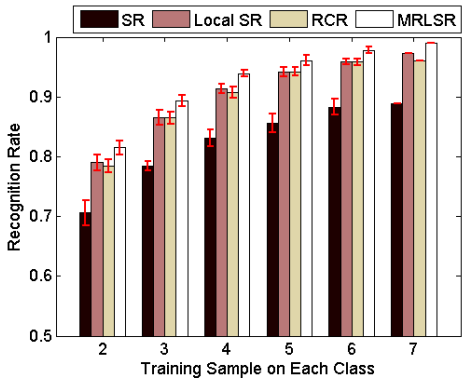

Fig. 3. Recognition rate comparisons with the SR and the Local SR with different training samples on the AR database without occlusion.

The higher recognition rate of our MRLSR on the AR database without occlusion indicates that manifold regularization is necessary for local sparse representation. Moreover, if the block size is small, local regularization of the MRLSR is more effective than global regularization of the RCR.

We also compare our method with the SR and the Local SR under different training samples N_{train}. The results are shown in Fig. 3. The training sample $N_{train} = 2$ means that for each person, we randomly choose two images from seven images for training. For the same training sample, we perform each recognition algorithm for 10 times.[4] As shown in Fig. 3, the recognition rates of MRLSR are higher than those of SR and Local SR.

Fig. 4 shows the comparison of the coding vectors. The top image is the absolute coding matrix $|\Lambda^T|$ obtained by the Local SR, while the bottom one is calculated by our MRLSR. From the two images, it can be seen that some coding values are suppressed by our method (see the coding values in the blue rectangle), while some coding values are enhanced (see the coding values in the red rectangle). As a result, our coding vectors hold more group sparsity [28], [29] than those of the Local SR. In this example, the true class of the testing image is 27. Some noise coding values cannot be suppressed by the Local SR. Hence, it misclassifies the testing image as 85.

Fig. 5 provides the convergence curves on the six testing face images. From this figure, we see that the energies gradually decrease as the iteration number increases, which experimentally verifies that the proposed homotopy-like algorithm is convergent.

B. Results on the AR Database With Occlusion

The second experiment is performed on the AR database with occlusion [34] to further evaluate the robustness of our MRLSR against occlusions. The selection of the training images is the same as the first experiment. Two separate

[2]The local SR method is similar to our method but without using the manifold regularization, that is, the coding matrix, Λ, is obtained by minimizing the equation of $\sum_{b=1}^{B}(1/2\,\|\mathbf{y}_b - \mathbf{X}_b\alpha_b\|_2^2 + \gamma_b\|\alpha_b\|_1)$.

[3]RSR is very robust to occlusion. Hence, we compared it on AR with occlusion dataset. TaSR is very efficient to illumination variations. Hence, we compared it on Extended Yale B and CMU-PIE datasets.

[4]Note that when the training sample N_{train} is set to 7, we perform all algorithms once, since we only have 7 training images for each person.

Fig. 4. Comparison of coding vectors. (a) Absolute coding matrix $|\Lambda^T|$ obtained by the Local SR. (b) Absolute coding matrix $|\Lambda^T|$ calculated by our MRLSR. (c) and (d) Close-ups of the coding matrices for the two classes, respectively. For each absolute coding matrix $|\Lambda^T|$, the bth row is the coding vector of the bth block. The true class of the testing image is 27. The Local SR misclassifies it as 85, while our MRLSR classifies it as 27 correctly.

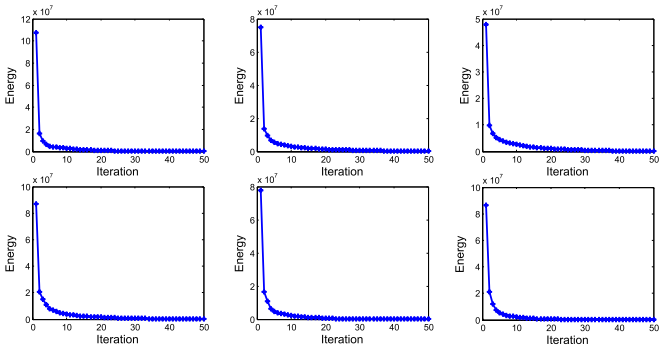

Fig. 5. Convergence curves on six testing face images.

TABLE II
RECOGNITION RATE COMPARISONS WITH THE RCR WITH DIFFERENT BLOCK SIZES ON THE AR DATABASE WITH OCCLUSION

Algorithm \ Block Size	12*16		12*8		8*16		8*8	
	Sunglass	Scarf	Sunglass	Scarf	Sunglass	Scarf	Sunglass	Scarf
Local SR	83.17%	88.67%	82.50%	88.00%	91.00%	90.33%	90.50%	90.33%
RCR	79.00%	89.00%	59.00%	80.67%	74.67%	89.17%	52.17%	81.17%
MRLSR	**85.83%**	**93.00%**	**86.50%**	**92.83%**	**92.83%**	**94.17%**	**91.67%**	**93.00%**

TABLE III
RECOGNITION RATE COMPARISONS WITH THE SR, LOCAL SR, RCR, AND RSR ON THE AR DATABASE WITH OCCLUSION

Algorithm	Rec. Rate		Time (second)
	Sunglass	Scarf	
Local SR	91.00%	90.33%	6.2
RCR	79.00%	89.17%	28.5
RSR	86.17%	86.00%	97.3
MRLSR	**92.83%**	**94.17%**	8.1

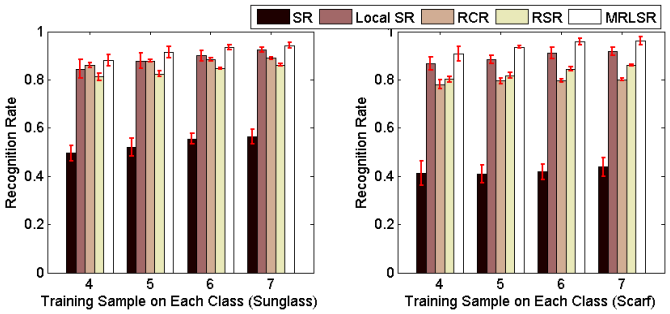

Fig. 6. Recognition rate comparisons with the SR and the Local SR with different training samples on the AR database with occlusion. The left sub-figure shows comparisons under Sunglass occlusions, while the right sub-figure gives comparisons under Scarf occlusions.

subsets with sunglass and scarf occlusions are utilized for testing. For each person, we select six images with sunglass occlusion and six images with scarf occlusion (Fig. 2).

The comparisons with the RCR method under different block sizes are shown in Table II. As shown in this table, the recognition rates of our MRLSR are higher than those of the RCR and the Local SR, especially when the block size is small. The recognition rates with the block size 8×16 are better than the recognition rates with the block size 12×8. Therefore, we conclude that when the image is occluded by disguise, the local methods could be sensitive to the image partition. Thanks to manifold regularization, our method is less sensitive to the image partition, as compared with the RCR method.

We also compare our MRLSR with the RSR, an efficient face recognition algorithm that is very robust against occlusions. The results, including recognition rate and running time, are shown in Table III. We select the best recognition rates in Table II as the results of Local SR, RCR, and MRLSR. As shown in Table III, our MRLSR is 12 times faster than RSR, while it still holds higher recognition rate.

The comparative results with different training samples are shown in Fig. 6. The selection of the training sample is the same as the first experiment. For each person, we select four testing images (two images with sunglass occlusion and two images with scarf occlusion). The testing images are reselected at each round of experiment. For the same training and testing images, we perform each recognition algorithm for 10 times. As shown in Fig. 6, the recognition rates are much higher than those of the SR, which are derived from the utilization of the local representation. Moreover, all the results of the MRLSR are higher than those of the Local SR.

Fig. 7. Image samples from the Extended Yale B database.

TABLE IV
RECOGNITION RATE COMPARISONS WITH THE RCR WITH DIFFERENT BLOCK SIZES ON THE EXTENDED YALE B DATABASE

Algorithm \ Block Size	16*14	16*8	8*14	8*8
Local SR	92.63%	93.61%	93.02%	94.74%
RCR	84.22%	78.86%	81.71%	73.55%
MRLSR	**95.09%**	**95.72%**	**95.58%**	**95.58%**

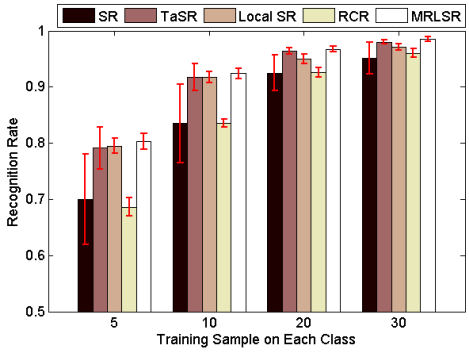

Fig. 8. Recognition rate comparisons with the SR, the TaSR, and the Local SR with different training samples on the Extended Yale B database.

C. Results on the Extended Yale B Database

The Extended Yale B database [35] consists of 16 128 face images of 38 subjects under 9 poses and 64 laboratory-controlled illumination conditions. Here, 2414 frontal images are selected. The resolution of each cropped gray-scale image is 64×56. Some examples are shown in Fig. 7. From this figure, we can see that it is even hard for people to distinguish or recognize some face images as they are corrupted by strong illumination changes.

Table IV provides comparative results with the RCR under different block sizes. For each subject, we randomly select 10 images for training, and use the rest for testing. As shown in this table, the recognition rates of our MRLSR are all at least 10% higher than other approaches, especially when the block size is small. This result indicates that when face images are strongly corrupted by illuminations, manifold regularization adopted in the MRLSR performs better than global regularization used in the RCR.

Fig. 8 shows experimental comparisons with different training samples. In this experiment, all images, including the training and testing images, are down-sampled to the size of 32×28. For each subject, we randomly select N_{train} ($N_{train} = 5, 10, 20, 30$) images for training, and use the rest for testing. For the same training and testing images, we perform each recognition algorithm for 10 times. As shown in Fig. 8, although the image resolution is very low, the proposed MRLSR algorithm can still provide high recognition rate.

TABLE V
RECOGNITION RATE COMPARISONS WITH THE RCR WITH DIFFERENT BLOCK SIZES ON THE CMU-PIE DATABASE

Algorithm \ Block Size	16*16	16*8	8*16	8*8
Local SR	95.21%	95.85%	95.28%	95.58%
RCR	95.96%	94.94%	93.51%	90.90%
MRLSR	**97.24%**	**97.36%**	**96.79%**	**96.68%**

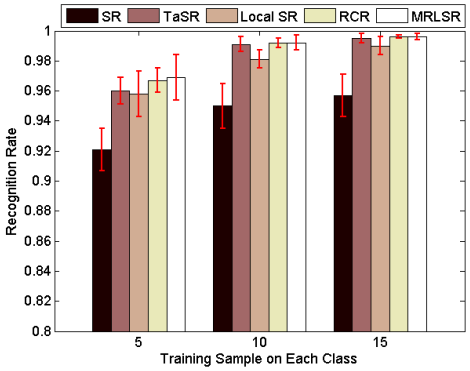

Fig. 9. Recognition rate comparisons with the SR, the TaSR, and the Local SR with different training samples on the CMU-PIE database.

TABLE VI
OUR RECOGNITION RATES UNDER DIFFERENT VALUES OF ϵ (BLOCK SIZE 8×8)

Database \ Parameter	10^{-1}	10^{-2}	10^{-3}	10^{-4}
AR without Occlusion	99.14%	99.14%	99.00%	99.00%
AR (sunglass occlusion)	91.50%	91.50%	91.67%	91.67%
AR (scarf occlusion)	92.67%	92.83%	93.00%	93.00%
YelaB	95.53%	95.58%	95.58%	95.58%

D. Results on the CMU-PIE Database

Similar comparative experiments are performed on the CMU-PIE database [36]. This face database contains images of 68 subjects under 13 poses, 43 illumination conditions, and 4 expressions. Only the frontal face images are selected. They are cropped and resized to 64×64 pixels.[5]

Table V shows experimental comparison with the RCR with different block sizes. For each subject, we only randomly select 5 images for training, and use the rest for testing. As shown in Table V, the MRLSR provides higher recognition rate.

Fig. 9 provides more comparisons with different training samples. Similar to the configuration for the Extended Yale B database, for each subject, we randomly select

[5]Downloaded from www.zjucadcg.cn/dengcai/data/data.html.

TABLE VII
OUR RECOGNITION RATES UNDER DIFFERENT VALUES OF λ (BLOCK SIZE 8×8)

Database \ Parameter	0.05	0.025	0.01	0.0025	0.001	0.000
AR without Occlusion	97.85%	98.28%	99.00%	99.00%	99.14%	95.72%
AR (sunglass)	91.67%	91.67%	91.67%	92.00%	92.00%	90.50%
AR (scarf)	92.00%	92.33%	93.00%	91.00%	91.00%	90.33%
YelaB	95.48%	95.53%	95.58%	95.58%	95.58%	94.74%

N_{train} ($N_{train} = 5, 10, 15$) images for training, and use the rest for testing. For the same training and testing images, we perform each recognition algorithm for 10 times. Again, from this figure, it can be seen that our algorithm provides the highest recognition rates.

E. Parameter Evaluation

In the above experiments, the parameter ϵ of Algorithm 1 is set to 10^{-3}. We evaluate the stability of the MRLSR under different choices of ϵ. The results are shown in Table VI. From this table, it can be seen that the recognition rates are very similar, which indicates that our algorithm is stable under variations of the sparse regularization parameter.

We further evaluate the stability of the MRLSR under different choices of λ, and the results are shown in Table VII. As shown in this table, our result is a little sensitive to the parameter λ, especially in AR with occlusion dataset. Fortunately, our results are all better than the corresponding results without using manifold regularization listed in the last column of Table VII ($\lambda = 0$). Hence, we can conclude that it is necessary to incorporate the manifold regularization.

VI. CONCLUSION

In this paper, we propose the manifold regularization for local sparse representation. All the coding vectors obtained by our MRLSR model can hold more group sparsity than those obtained by the Local SR. As a consequence, our face recognition rates are higher. In the future, we hope to incorporate other terms into our MRLSR model to enhance its group sparsity. As a general sparse representation method, the proposed MRLSR can also be applied to other classification-based applications, such as language classification and character recognition. Moreover, manifold regularization can be utilized in other local models.

REFERENCES

[1] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition*, 2nd ed. New York, NY, USA: Springer-Verlag, 2011.

[2] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[3] W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 36, no. 2, pp. 458–466, Apr. 2006.

[4] H. Zhao and P. C. Yuen, "Incremental linear discriminant analysis for face recognition," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 38, no. 1, pp. 210–221, Feb. 2008.

[5] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 989–993, Jul. 2008.

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

[7] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color face recognition for degraded face images," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 39, no. 5, pp. 1217–1230, Oct. 2009.

[8] S. Du and R. K. Ward, "Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1165–1175, Sep. 2010.

[9] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 625–632.

[10] Y. W. Wong, K. P. Seng, and L.-M. Ang, "Radial basis function neural network with incremental learning for face recognition," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 940–949, Aug. 2011.

[11] C.-P. Chen and C.-S. Chen, "Intrinsic illumination subspace for lighting insensitive face recognition," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 422–433, Apr. 2012.

[12] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[13] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2224–2231.

[14] L. Wang, H. Wu, and C. Pan, "Tone-aware sparse representation for face recognition," in *Proc. IEEE Conf. Autom. Face Gesture Recognit.*, Apr. 2013, pp. 1–6.

[15] J. Lu, Y.-P. Tan, G. Wang, and G. Yang, "Image-to-set face recognition using locality repulsion projections and sparse reconstruction-based similarity measure," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1070–1080, Jun. 2013.

[16] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[17] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[18] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[19] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 846–853.

[20] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[21] Q. Liu, H. Lu, and S. Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004.

[22] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[23] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, no. 6, pp. 1233–1258, 1987.

[24] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of Gabor magnitude and phase for face recognition," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1349–1361, May 2010.

[25] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 448–461.

[26] X. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 3493–3500.

[27] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Advances in Neural Information Processing Systems 23*. Red Hook, NY, USA: Curran Associates, Inc., 2010, pp. 1813–1821.

[28] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1873–1879.

[29] Y.-W. Chao, Y.-R. Yeh, Y.-W. Chen, Y.-J. Lee, and Y.-C. F. Wang, "Locality-constrained group sparse representation for robust face recognition," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 761–764.
[30] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems*, vol. 14. Cambridge, MA, USA: MIT Press, 2002, pp. 585–591.
[31] S. Shan, X. Chen, and W. Gao, "Face recognition based on non-corresponding region matching," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 1060–1067.
[32] D. L. Donoho and Y. Tsaig, "Fast solution of l_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
[33] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[34] A. Martinez and R. Benavente, "The AR face database," CVC, New Delhi, India, Tech. Rep. 24, Jun. 1998.
[35] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.
[36] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 53–58.

Lingfeng Wang received the B.S. degree in computer science from Wuhan University, Wuhan, China, in 2007. He is currently working toward the Ph.D. degree with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision and image processing.

Huaiyu Wu received the B.E. and M.E. degrees from Beijing University of Aeronautics and Astronautics, Beijing, China, in 2000 and 2003, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from Chinese Academy of Sciences, Beijing, in 2008.

He was a Post-Doctoral Lecturer with the School of Electrical Engineering and Computer Science, Peking University, Beijing, from 2008 to 2011. He is currently an Associate Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing. His research interests include 3-D computer vision, visual shape perception and analysis, and interactive computer graphics.

Chunhong Pan received the B.S. degree in automatic control from Tsinghua University, Beijing, China, in 1987; the M.S. degree from Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 1990; and the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2000.

He is a Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, image processing, computer graphics, and remote sensing.