# Face Recognition by Exploiting Local Gabor Features With Multitask Adaptive Sparse Representation

Leyuan Fang, *Member, IEEE*, and Shutao Li, *Senior Member, IEEE*

***Abstract*— For a human face, the Gabor transform can extract its multiple scale and orientation features that are very useful for the recognition. In this paper, the Gabor-feature-based face recognition is formulated as a multitask sparse representation model, in which the sparse coding of each Gabor feature is regarded as a task. To effectively exploit the *complementary yet correlated* information among different tasks, a flexible representation algorithm termed multitask adaptive sparse representation (MASR) is proposed. The MASR algorithm not only restricts Gabor features of one test sample to be jointly represented by training atoms from the same class but also promotes the selected atoms for these features to be varied within each class, thus allowing better representation. In addition, to use the local information, we operate the MASR on local regions of Gabor features. Then, by considering the structural characteristics of the face and the effects of the external interferences, a structural–residual weighting strategy is proposed to adaptively fuse the decision of each region. Experiments on various datasets verify the effectiveness of the proposed method in dealing with face occlusion, corruption, small number of training samples, as well as variations of lighting and expression.**

***Index Terms*— Decision fusion, face recognition, Gabor wavelet, multitask sparse representation, sparse residual.**

## I. Introduction

FACE recognition has attracted much attention due to its wide range of potential applications, such as biometric, identity authentication, law enforcement, and video surveillance [1]. Even though human eye can detect and identify faces in a scene easily, building an automatic system that accomplishes such a task is still very challenging. The challenges are mainly from large variations of illumination, expression, pose, corruption, and occlusion appeared in real face [2]–[4].

Up to now, many researchers have resorted to the design of effective face representation techniques, which aim to extract robust and discriminant features. The feature extraction approaches are mainly divided into two categories: 1) holistic feature based [5]–[7] and 2) local feature based [8]–[10]. Holistic-feature-based approaches, including the principal component analysis (PCA) [5], linear discriminant analysis (LDA) [7], and independent component analysis [6], have been popularly used for face recognition. Nonetheless, more recent works [8]–[12] attempt to develop local-feature-based techniques, since they are more stable to local changes, such as illumination, expression, and occlusion [13]–[15]. One of the most successful local descriptors is the Gabor wavelet [16], [17], due to its biological relevance and computational properties [18]. The Gabor wavelet whose kernels are similar to receptive field of the human cortical simple cells could effectively extract the image local orientation features at multiple scales. Furthermore, recent biological studies [19] have revealed that the scale and orientation information of retinal is very important for visual perception. However, effectively utilizing such multiscale-orientation information is a nontrivial work.

On the other hand, many efforts have been made to develop the discriminative face classification techniques. Typical classifiers include the nearest neighbor, neural network [20], and support vector machines [21], among others. Recently, Hinton and Salakhutdinov [22] proposed a deep learning (DL) method to train a deep neural network, which is also successfully applied for the face recognition task [23], [24]. Since the trained deep network can achieve a better representation of the face data, the DL-based face recognition methods obtain very good performances. However, these methods always require a very huge amount of training data (e.g., 4.4 million training faces in [24]) and computation cost to tune the deep network. Inspired by the sparse coding mechanism of human vision system [25]–[28], one more promising direction is to incorporate the *sparsity priori* into the classification scheme [29], [30]. Wright *et al.* [29] introduced a sparse representation-based classification (SRC) method, which represents a test sample using a small number of training samples (atoms) adaptively chosen out of all the candidate training samples (a structural dictionary). The classification result is then determined by evaluating which class of training samples has the least representation residual. In the original SRC, Wright *et al.* [29] only adopted the simple holistic descriptors (such as PCA and LDA) to reduce the dimensionality of the training and test samples.

Manuscript received May 28, 2014; revised January 21, 2015; accepted March 6, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61172161, in part by the National Natural Science Foundation for Distinguished Young Scholars of China under Grant 61325007, and in part by the Fundamental Research Funds for the Central Universities, Hunan University, Changsha, China. The Associate Editor coordinating the review process was Dr. Zheng Liu.

The authors are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: fangleyuan@gmail.com; shutao_li@hnu.edu.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/TIM.2015.2427893

Yang *et al.* [31] proposed a statistical local-feature-based robust kernel representation (SLF-RKR) method, which combines the SRC with the local binary projection (LBP) features and obtains better performance than the original SRC method. Yang and Zhang [32] proposed a Gabor-feature-based SRC (GSRC), which incorporates the Gabor local features into SRC. Since the original SRC can only process one test data at a time with its vector pattern, GSRC simply concatenated Gabor features of all scales and orientations into a long vector. Though such an approach provides promising recognition results, the simple features concatenating strategy might not fully exploit the Gabor information of different scales and orientations.

In this paper, to utilize the Gabor multiscale-orientation information, we first formulate the multifeature classification problem as a multitask sparse representation model [33]–[35]. More specifically, assuming that the Gabor wavelet creates $o \times s$ types of scale and orientation features for one face image, we can generate $o \times s$ representation tasks in which the sparse representation on each type of feature is regarded as one task. To exploit the correlations among different features, a multitask joint sparse representation (MJSR) model [35] can be used to confine all these features to share the same sparsity pattern via a joint sparse constraint. However, such a constraint is too restrictive, since different kinds of features even associated with the same subject might exist large differences. In other words, to achieve a more effective representation, each type of feature should have a certain degree of its own *freedom* on the sparsity. Based on this idea, we further propose a flexible representation algorithm called multitask adaptive sparse representation (MASR). More specifically, since different features all correspond to the same subject, the MASR still has the same sparsity pattern at class-level, which enforces all the features to choose the training atoms from the same class. Moreover, within each class, motivated by [36]–[39], the MASR adopts a flexible atom selection scheme that allows different features to adaptively choose their own appropriate atoms. In this way, the MASR not only combines the information of the different features (tasks) for discrimination but also achieves a more effective representation for each specific type of feature.

As Gabor wavelet only focuses on local areas, following [8], [13], and [14], we partition the Gabor feature image into a number of small regions and perform the MASR on each region. Intuitively, different regions in the face image should have different contributions for the final classification [13], [14], [40]. Therefore, a structural–residual region weighting (SRW) strategy is proposed to adaptively assign the weight for each region. Specifically, the SRW strategy first considers the structural characteristics of the human face, which can assign large weights for distinguished structural regions (e.g., nose and eye) and small weights for smooth regions (e.g., cheek and forehead). In addition, the sparse coding residual is utilized on the SRW strategy to reduce the effects of the external interferences (e.g., expression and occlusion).

The rest of this paper is organized as follows. In Section II, we briefly review the SRC and Gabor wavelet transform. Section III introduces the proposed MASR method to exploit the Gabor multiscale-orientation information and the SRW strategy to fuse the recognition result of each local region. The experimental results are presented in Section IV. Section V concludes this paper and suggests future works.

## II. Related Works

### A. Sparse Representation-Based Classification

Denote by $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_c, \ldots, \mathbf{A}_C] \in \mathrm{R}^{P\times N}$ the structural dictionary formed by the whole training samples, where $\mathbf{A}_c = [\mathbf{a}_c^1, \mathbf{a}_c^2, \ldots, \mathbf{a}_c^{N_c}] \in \mathrm{R}^{P\times N_c}$ is the subdictionary with samples (atoms) from class $c$, $C$ is the number of classes, $N_c$ is the number of atoms in subdictionary $\mathbf{A}_c$ and $N = \sum_{c=1}^{C} N_c$ is the total number of atoms in $\mathbf{A}$. The SRC has the basic assumption [29] that a test sample $\mathbf{y}_c \in \mathrm{R}^{P\times 1}$ from class $c$ can only be well represented by the subdictionary $\mathbf{A}_c$ with training samples (atoms) from the same class

$$\mathbf{y}_c = \mathbf{a}_c^1\mathbf{x}_c^1 + \mathbf{a}_c^2\mathbf{x}_c^2 + \cdots + \mathbf{a}_c^{N_c}\mathbf{x}_c^{N_c} = \mathbf{A}_c\mathbf{x}_c \tag{1}$$

where $\mathbf{x}_c = [\mathbf{x}_c^1, \mathbf{x}_c^2, \ldots, \mathbf{x}_c^{N_c}]$ is the coefficient vector for weighting each atom in class $c$. As the class label of the test sample $\mathbf{y} \in \mathrm{R}^{P\times 1}$ is unknown, the SRC sparsely represents the $\mathbf{y}$ utilizing the whole training dictionary $\mathbf{A}$ and pursuits the sparse coefficients $\hat{\mathbf{x}} \in \mathrm{R}^{N\times 1}$ as the error constrained form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \le \varepsilon \tag{2}$$

or the sparsity constrained form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t. } \|\mathbf{x}\|_0 \le K \tag{3}$$

where $\varepsilon$ is the error tolerance and $K$ is the sparsity level representing the number of selected atoms in the dictionary (also corresponding to the nonzeros coefficients in $\hat{\mathbf{x}}$). Greedy pursuit [41]–[44] and convex relaxation [45] are two types of typical algorithms for addressing the problems (2) and (3). The most commonly used greedy algorithm is the orthogonal matching pursuit (OMP) [41], which includes the following general steps at each iteration.

1) Select the best matched atom $\mathbf{a}_{\hat{j}}$ that exhibits the highest correlation with the residual vector $\mathbf{R} = \mathbf{y} - \mathbf{A}\mathbf{x}$ by searching over the whole dictionary

$$\mathbf{a}_{\hat{j}} = \arg\max_{\mathbf{a}_j} \left|\mathbf{a}_j^T\mathbf{R}\right|, \quad j = 1, \ldots, N. \tag{4}$$

2) Merge the newly selected atom's index with the previously selected atom's index set $\mathbf{J}$

$$\mathbf{J} = \mathbf{J} \cup \hat{j}. \tag{5}$$

3) Estimate sparse coefficient $\mathbf{x}$ by projecting the test subject $\mathbf{y}$ on $\mathbf{A}_\mathbf{J}$

$$\hat{\mathbf{x}} = \left(\mathbf{A}_\mathbf{J}^T\mathbf{A}_\mathbf{J}\right)^{-1}\mathbf{A}_\mathbf{J}^T\mathbf{y} \tag{6}$$

where the subdictionary $\mathbf{A}_\mathbf{J}$ is constructed by selected atoms. Note that the $\ell_2$ norm of the representation fidelity $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ in (2) can be changed to the regularized robust coding (RRC) [46] to obtain the better sparse vector $\mathbf{x}$. After the sparse coefficient vector $\hat{\mathbf{x}}$ is obtained, SRC determines the class label of the test

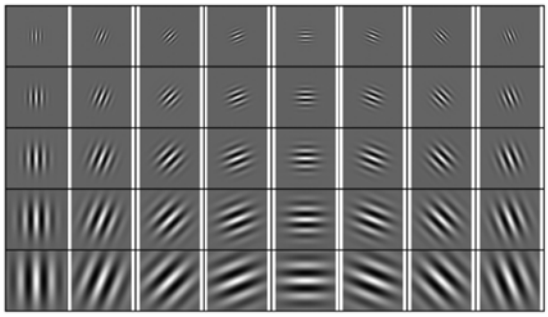


Fig. 1. Gabor wavelet with five scales and eight orientations.

sample $\mathbf{y}$ based on the minimum representation error criteria

$$\hat{c} = \mathrm{SRC}(\mathbf{y}) = \arg\min_c \|\mathbf{y} - \hat{\mathbf{y}}_c\|_2 = \arg\min_c \|\mathbf{y} - \mathbf{A}\Phi_c(\hat{\mathbf{x}})\|_2 \quad (7)$$

where $\Phi_c(\hat{\mathbf{X}})$ is an vector operator that preserves coefficients of $\hat{\mathbf{x}}$ corresponding to class $c$ and sets all other coefficients to zero.

### B. Gabor Wavelet

Gabor wavelet filters are defined as follows [10]:

$$\psi_{o,s}(z) = \frac{\|k_{o,s}\|^2}{\sigma^2} \cdot \exp\left(-\|k_{o,s}\|^2\|z\|^2/2\sigma^2\right) \cdot [\exp(ik_{o,s}z) - \exp(-\sigma^2/2)] \quad (8)$$

where $o$ and $s$ separately represent the orientation and scale of the Gabor filters, $z = (x, y)$ denotes the pixel, and $\sigma$ is the ratio of the Gaussian window width to wavelength. The wave vector $k_{o,s}$ is defined as

$$k_{o,s} = k_s \cdot \exp(i\phi_o) \quad (9)$$

where $k_s = k_{\max}/f_s$ and $\phi_o = (\pi \cdot o)/8$. $k_{\max}$ is the maximum frequency and $f_s$ is the spacing factor between kernels in the frequency domain.

Gabor wavelet can have a number of different types by altering the scale $S$ and orientation $O$. Fig. 1 shows the Gabor wavelets with five scales and eight orientations. As can be observed, the Gabor wavelet reflects various kinds of edge and bar details with different orientations and takes abundant frequency information with different scales. Therefore, the Gabor wavelet can extract more details in some meaningful local regions of face (e.g., nose, eyes, and mouth), which are very useful for recognition. The convolution of an input face image with the Gabor wavelet creates $O \times S$ magnitude images and $O \times S$ phase images. Since the magnitude information contains the variation of local energy, this paper only selects magnitude images as the Gabor features. In addition, to more effectively utilize the Gabor local information [8], [13], [14], we will partition the feature image into a set of local regions.

## III. Proposed Method for Local Gabor-Feature-Based Face Recognition

### A. Multitask Sparse Representation Model for Gabor Features

Suppose we have a training dictionary $\mathbf{A}$ and a test sample $\mathbf{y}$, as described aforementioned, the Gabor wavelet can generate $O \times S$ feature dictionaries $\{\mathbf{A}^1, \ldots, \mathbf{A}^{O\times S}\}$ and test samples $\{\mathbf{y}^1, \ldots, \mathbf{y}^{O\times S}\}$ of different orientations and scales. We arrange all test samples and atoms of training dictionaries into column vectors and denote their sparse coefficients as a matrix $\mathbf{M} = [\mathbf{x}^1, \ldots, \mathbf{x}^{O\times S}] \in \mathrm{R}^{N\times(O\times S)}$, where $N$ stands for the dimension of each sparse coefficient (corresponding to the number of atoms in the dictionary $\mathbf{A}$). The sparse matrix can also be represented as row vectors $\mathbf{M} = [\mathbf{x}_1; \ldots; \mathbf{x}_j; \ldots; \mathbf{x}_P]$, where $\mathbf{x}_j$ is one row of the matrix $\mathbf{M}$. Note that, for simplicity, we here only utilize the whole image as the feature. In the later section, we will discuss how to partition the image into regions and utilize the SRW strategy to fuse the result of each region.

After Gabor multiscale-orientation dictionaries $\{\mathbf{A}^1, \ldots, \mathbf{A}^{O\times S}\}$ and test samples $\{\mathbf{y}^1, \ldots, \mathbf{y}^{O\times S}\}$ are obtained, we aim to utilize the information from different scales and orientations to make a single decision for the recognition. Based on the SRC model, one simple way is to rewrite (3) into a multitask sparse representation problem (Fig. 2)

$$\{\hat{\mathbf{x}}^i\}_{i=1}^{O\times S} = \arg\min_{\{\mathbf{x}^i\}} \sum_{i=1}^{o\times s} \|\mathbf{y}^i - \mathbf{A}^i\mathbf{x}^i\|_2 \quad \text{s.t. } \|\mathbf{x}^i\|_0 \le K \quad \forall 1 \le i \le (O \times S). \quad (10)$$

However, this formulation separately pursuits the sparse coefficient $\hat{\mathbf{x}}^i$ for each task and thus does not consider the relationship among the different tasks (orientations and scales). To combine the information among the Gabor orientations and scales, we can use a joint sparse assumption [34], [35] that the sparse coefficients of different tasks have the same sparse pattern. That is, for different tasks, the positions of the nonzero coefficients in all the sparse vectors $\mathbf{x}^1, \ldots, \mathbf{x}^{O\times S}$ are identical, while coefficient values might be varied. Under this assumption, the nonzero coefficients in $\mathbf{M}$ should be on the same rows and a joint sparse regularization $\ell_{\mathrm{row},0}$ can be placed on the $\mathbf{M}$ to select a small number of the representative nonzero rows

$$\|\mathbf{M}\|_{\mathrm{row},0} = \|[\|\mathbf{x}_1\|_2; \ldots; \|\mathbf{x}_j\|_2; \ldots; \|\mathbf{x}_P\|_2]\|_0 \quad (11)$$

where $\mathbf{x}_j$ is one row in the sparse coefficient matrix $\mathbf{M}$.

Then (10) can be rewritten as

$$\hat{\mathbf{M}} = \arg\min_{\{\mathbf{M}\}} \sum_{i=1}^{O\times S} \|\mathbf{y}^i - \mathbf{A}^i\mathbf{x}^i\|_2 \quad \text{s.t. } \|\mathbf{M}\|_{\mathrm{row},0} \le K. \quad (12)$$

In [34] and [35], the application of the joint sparse prior for multitask problem is termed as the MJSR model.

### B. Multitask Adaptive Sparse Representation

The MJSR model explained above can be directly used to exploit the Gabor information of different scales and orientations by enforcing the same sparsity pattern among them.

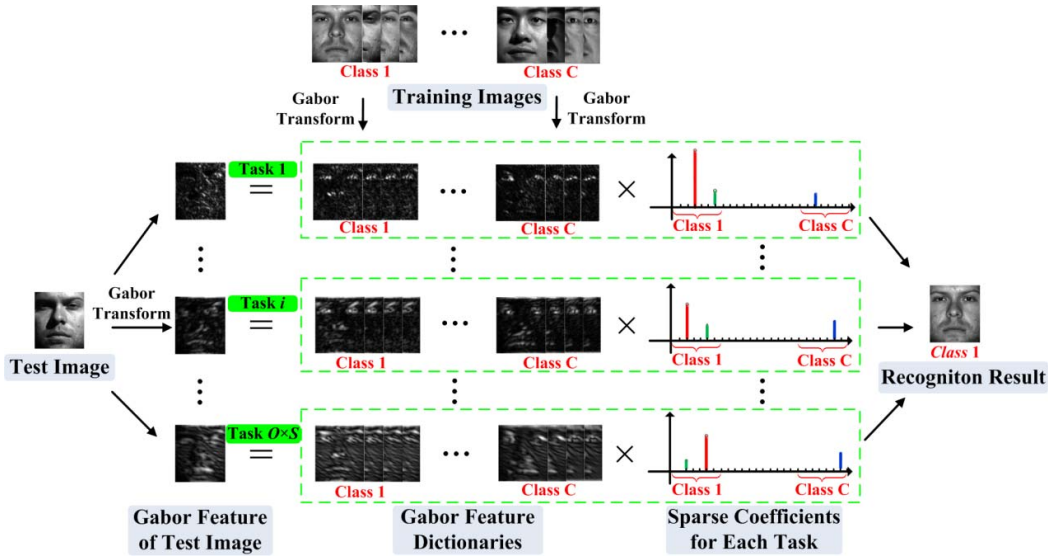



Fig. 2. Schematic of the multitask sparse representation model for the Gabor multiscale-orientation feature-based face recognition.

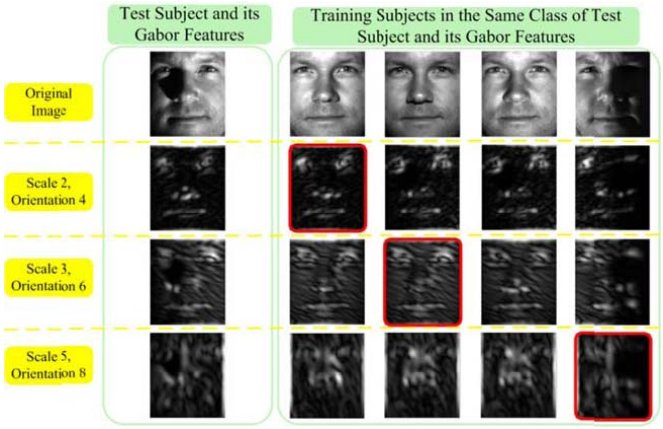



Fig. 3. Multiorientation-scale features have different sparsity pattern within each class for the better representation. In each type of feature, the best representation atom for the test sample is added with a red border.

However, assuming that all the orientation and scale features have the same sparsity pattern might be too restrictive, since multiorientation-scale features are very different. That is, each kind of feature should have its own *freedom* in selecting the position of the atoms for a better representation. On the other hand, since all the multiscale-orientation features belong to the same subject, the *freedom* should be restricted within each class. Fig. 3 provides an example that indicates that within the same class, each type of feature selects atoms from different positions for a better representation. Therefore, the desired sparse coefficients for multiscale-orientation features should favor the same class-level sparsity pattern, while allowing distinct sparsity pattern within each class. According to this idea, we propose a flexible model called MASR with an adaptive atom selection strategy. Similar to [36]–[39], the adaptive atom selection is achieved by the adoption of the adaptive set, in which each adaptive set, $d_h \in \mathrm{R}^{(o\times s)}$, $h = 1, 2, \ldots,$ can be denoted as the indices of a set of nonzero coefficients corresponding to the same class in the sparse coefficient matrix $\mathbf{M}$. In each column (task $i$) of $\mathbf{M}$, each adaptive set $d_h$ contains only one index, which represents the position of the selected atom for the task $i$. Indices in one $d_h$ are not restricted to be the same. That is, different features of a test sample have a certain degree of their own freedom to select varied atoms within each class. In this way, the positions of nonzero coefficients for different tasks can be varied, but still be enforced in the same class. Note that if all indices in $d_h$ are the same, this can be regarded as a special case, which is the same as that in the MJSR model. By combining the adaptive set with the $\ell_{\mathrm{row},0}$ norm, we create a novel adaptive sparse norm $\ell_{\mathrm{adaptive},0}$ on $\mathbf{M}$, which can select a small number of the representative adaptive sets from $\mathbf{M}$

$$\|\mathbf{M}\|_{\mathrm{adaptive},0} = \|[\|\mathbf{x}_{d_1}\|_2; \|\mathbf{x}_{d_2}\|_2; \ldots]\|_0 \tag{13}$$

where $\mathbf{x}_{d_h}$ denotes the vector composed of coefficients associated with the $h$th adaptive set $d_h$. Since the adaptive sets are flexible in selecting atoms of different positions within each class, the $\ell_2$-norm applied over each adaptive set enables each feature (at a specific orientation and scale) to adaptively select its appropriate atoms and thus will provide a better representation. By replacing the $\ell_{\mathrm{row},0}$ regularization in (12) with the adaptive sparse term $\ell_{\mathrm{adaptive},0}$, the MASR model can be defined as

$$\hat{\mathbf{M}} = \arg\min_{\{\mathbf{M}\}} \sum_{i=1}^{O\times S} \|\mathbf{y}^i - \mathbf{A}^i\mathbf{x}^i\|_2 \quad \text{s.t. } \|\mathbf{M}\|_{\mathrm{adaptive},0} \le K. \tag{14}$$

To solve this problem, we introduce a MASR solution algorithm based on the modification of the OMP algorithm [41]. As mentioned in the above section, the OMP can only select one atom at each iteration and is designed for separately solving each task. To meet the requirement of the MASR model, one major modification is to use an adaptive set selection strategy. For each iteration

of this strategy, it will select a new representative adaptive set with the following steps: 1) find the best representation atom for each class and each task; 2) combine the best atoms across the tasks for each class into a cluster; and 3) select the best cluster and recorder the atoms' indices of that cluster as one selected adaptive set. In addition, motived by the strength of multiselection strategy [42], [47], several adaptive sets (e.g., $2K$) are allowed to be selected and then a refining procedure is adopted to prune part of them (e.g., $K$) at each iteration. The details of the MASR solution algorithm are described in Fig. 4. In Fig. 4, the $\mathbf{J}$ is denoted as the index matrix, in which each row contains one adaptive set.

After obtaining sparse coefficients $\hat{\mathbf{x}}^1, \ldots, \hat{\mathbf{x}}^{O\times S}$ for all the tasks (Gabor features), we make a single decision on the class label for multitask test samples $\mathbf{y}^1, \ldots, \mathbf{y}^{O\times S}$ simultaneously based on the lowest total representation error accumulated over all the tasks

$$\hat{c} = \arg\min_{c} \sum_{i=1}^{O\times S} \left\| \mathbf{y}^i - \mathbf{A}^i \Phi_c^i(\hat{\mathbf{x}}^i) \right\|_2 \tag{15}$$

where $\Phi_c^i(\hat{\mathbf{x}}^i)$ keeps the coefficients of $\hat{\mathbf{x}}^i$ corresponding to $c$th class while setting all other coefficients to zero.

### C. Adaptive Region Fusion With Structural–Residual Weighting

As described in [14], the Gabor wavelet focuses on the local information. To more effectively utilize the locality information, the Gabor feature images are partitioned into $T$ nonoverlapping local regions of size $p \times q$. Then, we perform (14) and (15) on each local region, in which one classification result $c_t$ can be obtained. Finally, these local results are combined to form the final decision for the test sample, which is formulated as

$$\hat{c} = \sum_{t=1}^{T} w_t \cdot c_t \tag{16}$$

where $w_t$ is the weight of $c_t$.

We may simply use the majority voting [48] to select the equal weight, $w_t = 1/T$, for each region. However, as different facial areas have different effects for recognition [13], [14], [40], the weights are needed to be adaptively assigned. One intuitive scheme is to assign the weights according to the structural characteristics of the face. That is, some distinguished structural regions (e.g., nose and eye) favor large weights, whereas smooth regions (e.g., cheek and forehead) deserve small weights. On the other hand, the effects of the external interferences (such as big variations of illumination, expression, and occlusion) should also be considered on weighting the regions. According to these ideas, we propose a SRW strategy. The SRW first calculates the recognition rate $c_t^{\text{Validation}}$ for each region on a validation dataset and adopts them as the structural weights $w_t^{\text{Struc}}, t = 1, \ldots, T$, [Fig. 5(a)]. This is because high recognition rates should correspond to distinguished structural regions while low recognition rates should belong to smooth regions, as shown in Fig. 5(a). Note that the validation datasets are selected from the training samples, which are assumed to be not affected by the corruption and occlusion, and thus the structural weights reflect only the structural information of the face. In addition, corrupted or occluded areas cannot be well represented by the training dictionary with a rather big residual error, whereas the facial area can be well reconstructed with respect to the training dictionary with a small residual error. Therefore, we can use the sparse residual error $\|\mathbf{R}_t\|_2$ to compute the residual weight $w_t^{\text{Resid}}$ for each region with the logistic function [49]

$$w_t^{\text{Resid}} = 1/(1 + 1/\exp(-\beta \cdot \|\mathbf{R}_t\|_2)), \quad t = 1, \ldots, T \tag{17}$$

where $\beta$ is a positive scalar. Since each task has one residual, $\mathbf{R}_t$ used in (17) is the mean of residuals from all the tasks. Finally, to consider both structural characteristics and external interferences in the test face, we create the structural–residual joint weight $w_t^{\text{Struc,Resid}}$ by fusing the above two weights

$$w_t^{\text{Struc,Resid}} = \left(w_t^{\text{Struc}} \cdot w_t^{\text{Resid}}\right) \Big/ \sum_{t=1}^{T} \left(w_t^{\text{Struc}} \cdot w_t^{\text{Resid}}\right). \tag{18}$$

Fig. 5 gives an example that shows the structural weights, residual weights, and the structural–residual weights for a test image from the AR dataset [50] with both the occlusion and illumination changes.

## IV. Experimental Results

The proposed method is termed the MASR-SRW, which performs the MASR model on each local region and then uses the SRW strategy to fuse the result of each region. In this section, the Extended Yale B [51], AR [50], and Multiple-Pose Illumination, and Expression (Multi-PIE) [52] datasets are used to test the performance of the proposed MASR-SRW method for dealing with face occlusion, corruption, small number of training samples, as well as variations of lighting and expression.

In these experiments, the proposed MASR-SRW approach is compared with the MASR-voting, MASR-No Partition (MASR-NoP), as well as other 11 competing face recognition algorithms: SRC-p [29], GSRC-NoP[1] [32], GSRC-p[1] [32], LBP[2] [9], local Gabor binary pattern histogram sequence (LGBPHS) [13], RRC[1] [46], SLF-RKR[1] [31], Volterrafaces[3] [48], MJSR-NoP [35], MJSR-voting [35], and MJSR-SRW [35] for comparison of results. The MASR-voting method uses the MASR model to obtain the recognition result of each local region and adopts the majority voting to assign equal weights for local regions, instead of the SRW strategy. The MJSR-voting and MJSR-SRW methods apply the MJSR model to get the recognition result of each region and then use the majority voting and SRW strategy to fuse the result of each region, respectively. In the MJSR model, its solution algorithm utilizes the greedy kind of algorithm as in

[1]The source code was downloaded at: http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm.

[2]The source code was downloaded at: http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab.

[3]The source code was downloaded at: http://people.seas.harvard.edu/~rkkumar/.

**MASR Solution Algorithm**

**Input:** $\left\{\mathbf{y}^i\right\}_{i=1}^{O\times S}$, multi-task test samples; $\left\{\mathbf{A}^i\right\}_{i=1}^{O\times S}$, multi-task dictionaries; $K$, sparsity level; $O$×$S$, number of tasks; $C$, number of classes; *Titer*, number of iteration.

**Output:** $\mathbf{M}$, sparse coefficients matrix.

**Initialization:** set the index matrix $\mathbf{J}=\varnothing$, residual vectors $\mathbf{R}^i=\mathbf{y}^i, i=1,...,O\times S$, the iteration counter $iter=1$.

**1:** Compute residual correlation vector $\mathbf{E}^{i,iter}$ for each task: $\mathbf{E}^{i,iter}=\left(\mathbf{A}^i\right)^T\mathbf{R}^{i,iter}, i=1,...,O\times S.$

**2:** Select 2$K$ adaptive sets by repeating the adaptive set selection strategy (step 2.1-2.4) 2$K$ times on $\left\{\mathbf{E}^{i,iter}\right\}_{i=1}^{O\times S}$.

Set iteration counter of the adaptive set selection strategy: *Diter*=1.

**2.1**: Find the best representation atoms' index $j_{c,Diter}^{i,iter}$ and the corresponding coefficient value $v_{c,Diter}^{i,iter}$ for each class and each task: $\left\{j_{c,Diter}^{i,iter}, v_{c,Diter}^{i,iter}\right\}\leftarrow\max\left|\mathbf{E}_c^{i,iter}\right|, i=1,...,o\times s; c=1,...,C,$ where $\mathbf{E}_c^{i,iter}$ represents the correlation coefficients in the $c$ class and $i$ task.

**2.2:** Combine the indexes $j_{c,Diter}^{i,iter}$ of the best atoms across the tasks for each class into a cluster $\mathbf{W}_{c,Diter}^{iter}$:

$$\mathbf{W}_{c,Diter}^{iter}\leftarrow\text{combine}\left\{j_{c,Diter}^{i,iter}\right\}_{i=1}^{O\times S},\ c=1,...,C,$$

where $\text{combine}\{\cdot\}$ denotes the operation that combines the indexes $j_{c,Diter}^{i,iter}$ of the best atoms across all the tasks ($i$=1,…,$O\times S$) together into a cluster. Then, sum the coefficients in that cluster: $\mathbf{V}_{c,Diter}^{iter}\leftarrow\sum_{i=1}^{O\times S}v_{c,Diter}^{i,iter}, c=1,...,C.$

**2.3:** Select the best cluster index $\hat{c}_{Diter}^{iter}$: $\hat{c}_{Diter}^{iter}\leftarrow\max_c\left(\mathbf{V}_{c,Diter}^{iter}\right), c=1,...,C$, and recorder the atoms' indexes of that cluster $\mathbf{W}_{c,Diter}^{iter}\left(\hat{c}_{Diter}^{iter}\right)$ as one selected adaptive set and reserve as a row in one temporary index matrix $\mathbf{J}_{temp}^{iter}$: $\mathbf{J}_{temp}^{iter}\left(Diter,:\right)=\mathbf{W}_{c,Diter}^{iter}\left(\hat{c}_{Diter}^{iter}\right)$. Then, set the associated coefficients (for the selected atoms) in $\mathbf{E}_c^{i,iter}$ to zero to make sure that the current selected atoms will not be chosen in the next iteration.

**2.4:** Check if $Diter\geq 2K$, stop the procedures and output the final temporary index matrix $\mathbf{J}_{temp}^{iter}$; otherwise, set *Diter* = *Diter*+1 and go to the step 2.1.

**3:** update the index matrix: $\mathbf{J}^{iter}=\mathbf{J}^{iter}\bigcup\mathbf{J}_{temp}^{iter}$.

**4.** Estimate the sparse representation coefficients $\mathbf{M}_{temp}^{iter}$: $\mathbf{M}_{temp}^{iter}\leftarrow\left(\left(\mathbf{A}_{\mathbf{J}^{iter}}^i\right)^T\mathbf{A}_{\mathbf{J}^{iter}}^i\right)^{-1}\left(\mathbf{A}_{\mathbf{J}^{iter}}^i\right)^T\mathbf{y}^i$, $i$=1,…,$O\times S$.

**5:** Prune $K$ adaptive sets by performing the adaptive set selection strategy (step 2.1-2.4) $K$ times on $\mathbf{M}_{temp}^{iter}$ and get the newly estimated index matrix $\mathbf{J}^{iter}$ (for more details about the pruning process, see [42]).

**6:** Update the sparse coefficient matrix $\mathbf{M}^{iter}$ with the newly estimated index matrix $\mathbf{J}^{iter}$ as the step 4 and compute the current residual: $\mathbf{R}^{i,iter}=\mathbf{y}^i-\mathbf{A}^i\mathbf{x}^{i,iter}, i=1,...,O\times S.$

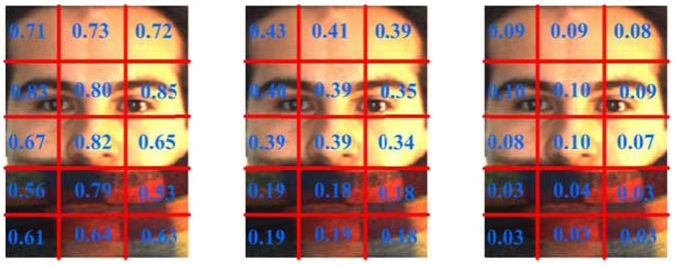
**7:** Check if $iter\geq Titer$, stop the procedures and output the final sparse coefficients matrix $\mathbf{M}$; otherwise , set *iter* = *iter*+1 and go to the step 1.

Fig. 4. MASR solution algorithm.

our MASR solution algorithm. To achieve fair comparison, the SRC-p and GSRC-p also partition image into a number of regions and use the majority voting to obtain the final result. The MASR-NoP, MJSR-NoP, and GSRC-NoP methods do not partition the image into regions and directly apply the MASR, MJSR, and GSRC algorithms on the each whole Gabor

TABLE I
RECOGNITION ACCURACY (%) WITH DIFFERENT PATCH SIZES FOR THE PROPOSED METHOD (MASR-SRW) ON EXTENDED YALE B, AR, AND CMU MULTI-PIE

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Extended Yale B | Patch Size | 34×34 | 32×32 | 30×30 | 28×28 | 26×26 | 24×24 | 22×22 |
| | Accuracy | 95.9 | 96.2 | 96.4 | **97.7** | 97.2 | 97.4 | 97.4 |
| AR | Patch Size | 28×28 | 26×26 | 24×24 | 22×22 | 20×20 | 18×18 | 16×16 |
| | Accuracy | 87.7 | 93.8 | 94.0 | 95.0 | **95.2** | **95.2** | **95.2** |
| CMU Multi-PIE | Patch Size | 20×20 | 18×18 | 16×16 | 14×14 | 12×12 | 10×10 | 8×8 |
| | Accuracy | 86.7 | 86.1 | 88.1 | 88.1 | 88.7 | **89.2** | **89.2** |




(a) (b) (c)

Fig. 5. Adaptive weights for a test image from the AR dataset [50] with both disguise and illumination changes (right part). (a) Structural weights. (b) Residual weights. (c) Normalized structural–residual weights.

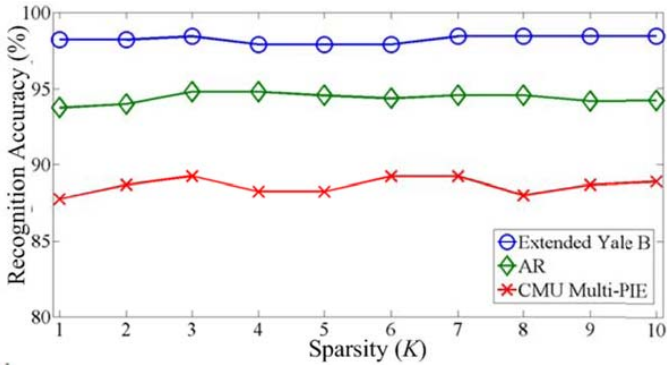



Fig. 6. Recognition accuracy (%) with different sparsity levels for the proposed method (MASR-SRW) on the extended Yale B, AR, and CMU multi-PIE.

feature image. In all the following experiments, the program of each approach has been run 10 times on each dataset and the average results are reported.

### A. Dataset Description

The Extended Yale B dataset [51] contains about 2414 frontal face images of 38 individuals. The images were captured under varying illumination conditions and normalized to 192 × 168. For each individual (class), there are about 64 images.

The AR dataset [50] consists of over 4000 frontal images from 126 individuals, including frontal views of faces with variations of facial expressions, illumination conditions, and occlusions. A subset that contains 50 males and 50 females is chosen for our test. For each individual, 26 pictures are taken in two separate sessions. In each session, the first seven images are simply with illumination and expression changes, while the remaining six images are with real occlusions (e.g., sunglasses and scarves). All images are normalized to 165 × 120.

The CMU Multi-PIE dataset [52] contains 68 human subjects with 41 368 face images. The images are taken under varying pose, illumination conditions, and expression. As in [53], our test chooses the frontal pose subset (C27), which consists of 3329 images in total. For each individual, there are about 49 images that are resized to 64 × 64.

### B. Parameters Setting

In the Gabor transform, the scale $S$ and orientation $O$ are set to five and eight, respectively. $\beta$ in (17) is empirically set to two. Since samples in the three test datasets are of different sizes, we need to select varied sizes of regions for different datasets. To examine the effect of the region size to our performance, we selected a number of different region sizes for the Extended Yale B, AR and CMU Multi-PIE datasets. For these three datasets, we randomly selected three samples per subject for training and another seven samples of this subject for testing. Table I shows the performance of the MASR-SRW method under different region sizes. As can be observed, as the region size decreases, the MASR-SRW method will have better results. When the region sizes are separately selected as 28 × 28, 20 × 20, and 10 × 10 for the Extended Yale B, AR, and CMU Multi-PIE datasets, respectively, the MASR-SRW method almost reaches the best performance. As smaller size (more regions) will increase the computational cost, region sizes 28 × 28, 20 × 20, and 10 × 10 are separately selected for the Extended Yale B, AR and CMU Multi-PIE in these experiments. Another parameter that might affect the recognition performance of the proposed MASR-SRW method is the sparsity level. We vary the sparsity level within the range $K \in \{1, 2, \ldots, 9, 10\}$ and examine the recognition rate of the MASR-SRW for each sparsity level on the Extended Yale B, AR, and CMU Multi-PIE, as shown in Fig. 6. As can be seen, the MASR-SRW algorithm is generally robust to various sparsity levels, since the accuracy difference is very small (within 1.5%). When the sparsity level $K = 3$, the MASR-SRW algorithm almost delivers the best results for these three datasets. Therefore, the sparsity level is chosen to be three in our experiments. The total iteration number *Titer* in our MASR solution algorithm is set to three. We note that utilizing larger iteration numbers may slightly enhance the performance while creating a more computational burden. For the structural weights learning, we select

TABLE II
RECOGNITION ACCURACY (%) ON THE EXTENDED YALE B DATASET WITHOUT OCCLUSION AND CORRUPTION

| Method | Number of Training Samples | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| LBP | 31.6 | 49.3 | 67.2 | 68.0 | 69.5 |
| LGBPHS | 43.6 | 62.7 | 76.4 | 78.9 | 80.3 |
| SRC-p | 52.1 | 80.3 | 87.8 | 92.3 | 93.6 |
| GSRC-NoP | 46.8 | 67.8 | 78.1 | 79.6 | 81.8 |
| GSRC-p | 70.0 | 88.9 | 92.1 | 95.2 | 95.2 |
| RRC | 41.7 | 59.7 | 78.6 | 80.1 | 82.1 |
| SLF-RKR | 42.3 | 69.8 | 85.3 | 90.4 | 92.1 |
| Volterrafaces | — | 87.2 | 89.3 | 93.6 | 95.1 |
| MJSR-NoP | 51.5 | 68.1 | 79.2 | 81.0 | 85.5 |
| MJSR-voting | 70.8 | 91.5 | 96.1 | 97.4 | 98.1 |
| MJSR-SRW | 72.2 | 91.9 | 96.4 | 97.8 | 98.1 |
| MASR-NoP | 55.0 | 73.8 | 84.5 | 85.8 | 88.1 |
| MASR-voting | 72.7 | 93.7 | 96.5 | 97.6 | **98.5** |
| MASR-SRW | **75.3** | **94.5** | **97.4** | **98.2** | **98.5** |

TABLE III
RECOGNITION ACCURACY (%) ON THE AR DATASET WITHOUT OCCLUSION AND CORRUPTION

| Method | Number of Training Samples | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| LBP | 74.4 | 76.6 | 77.0 | 82.3 | 86.8 |
| LGBPHS | 85.8 | 89.4 | 91.0 | 93.8 | 94.6 |
| SRC-p | 74.0 | 79.8 | 84.4 | 91.4 | 92.8 |
| GSRC-NoP | 76.4 | 79.4 | 80.6 | 85.6 | 89.4 |
| GSRC-p | 84.2 | 89.6 | 92.2 | 98.0 | 98.8 |
| RRC | 71.8 | 73.4 | 78.1 | 82.1 | 90.4 |
| SLF-RKR | **88.8** | 92.9 | 94.2 | 97.6 | 98.2 |
| Volterrafaces | — | 74.0 | 79.0 | 86.0 | 87.1 |
| MJSR-NoP | 76.8 | 80.6 | 81.8 | 86.2 | 89.3 |
| MJSR-voting | 86.3 | 92.4 | 94.2 | 99.4 | 99.4 |
| MJSR-SRW | 87.1 | 93.0 | 94.0 | 99.4 | 99.4 |
| MASR-NoP | 77.8 | 82.0 | 84.3 | 88.7 | 91.8 |
| MASR-voting | 87.6 | 93.2 | **94.8** | **99.8** | **99.8** |
| MASR-SRW | 88.2 | **93.6** | **94.8** | **99.8** | **99.8** |

one image from the training set as the test image and another two images as the training images. If the number of samples per subject in the training set are smaller than three, we only adopt the residual weight. The region sizes of the SRC-p, GSRC-p, MJSR-voting, and MJSR-SRW are set the same as those in our method. Other parameters in the compared methods are selected to reach their best results.

### C. Face Recognition Without Occlusion and Corruption

We first test the MASR-SRW, MASR-voting, MASR-NoP and other compared approaches on Extended Yale B, AR, and CMU Multi-PIE datasets with variations such as illumination and expression, but without occlusion and corruption. To make the face recognition problem more challenging, we adopt the small number of training sets. More specifically, in the Extended Yale B dataset, we randomly choose one–five samples per subject for training and 10 samples for testing.

TABLE IV
RECOGNITION ACCURACY (%) ON THE CMU MULTI-PIE DATASET WITHOUT OCCLUSION AND CORRUPTION

| Method | Number of Training Samples | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| LBP | 33.2 | 49.7 | 62.7 | 72.0 | 74.7 |
| LGBPHS | 55.1 | 71.1 | 85.2 | 88.9 | 91.4 |
| SRC-p | 57.5 | 73.6 | 88.6 | 91.3 | 92.7 |
| GSRC-NoP | 41.1 | 60.8 | 73.0 | 81.3 | 83.3 |
| GSRC-p | 60.2 | 76.6 | 87.0 | 92.6 | 93.5 |
| RRC | 56.6 | 78.8 | 87.6 | 91.2 | 94.6 |
| SLF-RKR | 57.7 | 76.7 | 88.7 | 92.6 | 95.5 |
| Volterrafaces | — | 79.7 | 90.6 | 92.5 | 95.8 |
| MJSR-NoP | 47.3 | 66.3 | 78.5 | 85.4 | 88.1 |
| MJSR-voting | 65.3 | 82.2 | 91.0 | 92.8 | 96.1 |
| MJSR-SRW | 65.8 | 82.5 | 91.0 | 92.5 | 96.1 |
| MASR-NoP | 48.7 | 68.9 | 81.8 | 87.7 | 90.5 |
| MASR-voting | **67.1** | **83.1** | 91.2 | **93.5** | 96.4 |
| MASR-SRW | **67.1** | 82.8 | **91.8** | **93.5** | **96.7** |

TABLE V
RECOGNITION ACCURACY (%) ON THE EXTENDED YALE B DATASET WITH SIMULATED OCCLUSION (FROM 10% TO 50%)

| Method | Experimental Setting A | | | | | Experimental Setting B | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Occlusion (%) | | | | | Occlusion (%) | | | | |
| | 10 | 20 | 30 | 40 | 50 | 10 | 20 | 30 | 40 | 50 |
| LBP | 93.0 | 88.7 | 77.2 | 72.0 | 57.1 | 65.0 | 63.4 | 59.4 | 46.8 | 28.6 |
| LGBPHS | 95.6 | 93.2 | 89.6 | 85.0 | 75.3 | 74.7 | 70.2 | 68.6 | 64.0 | 59.4 |
| SRC-p | **100** | **100** | 99.3 | 97.4 | 91.9 | 86.0 | 85.7 | 82.3 | 76.0 | 68.1 |
| GSRC-NoP | **100** | **100** | 99.8 | 96.5 | 87.4 | 77.7 | 77.0 | 74.5 | 67.6 | 60.0 |
| GSRC-p | **100** | **100** | 99.6 | 97.8 | 94.6 | 91.0 | 90.0 | 87.6 | 83.1 | 73.9 |
| RRC | **100** | **100** | 99.8 | 96.7 | 87.4 | 68.6 | 68.4 | 66.5 | 64.2 | 61.0 |
| SLF-RKR | **100** | **99.6** | 99.6 | 99.6 | 96.7 | 82.1 | 80.5 | 78.6 | 73.6 | 63.8 |
| Volterrafaces | **100** | 96.3 | 91.2 | 86.5 | 81.3 | 88.8 | 87.6 | 84.7 | 76.8 | 66.2 |
| MJSR-NoP | **100** | **100** | 99.3 | 97.4 | 89.8 | 79.1 | 78.4 | 76.3 | 71.4 | 65.9 |
| MJSR-voting | **100** | **100** | 99.8 | 99.3 | 95.2 | 94.5 | 92.8 | 91.6 | 89.8 | 79.0 |
| MJSR-SRW | **100** | **100** | **100** | 99.6 | 96.3 | 94.8 | 94.2 | 93.6 | 91.3 | 81.3 |
| MASR-NoP | **100** | **100** | 99.6 | 96.2 | 91.9 | 84.2 | 84.0 | 80.8 | 74.3 | 70.9 |
| MASR-voting | **100** | **100** | **100** | 99.6 | 96.9 | 95.7 | 95.3 | 93.7 | 90.2 | 81.1 |
| MASR-SRW | **100** | **100** | **100** | **100** | **97.8** | **96.3** | **95.7** | **95.0** | **92.3** | **83.9** |

In the AR dataset, for each subject, one–five samples are randomly selected from the first seven images of session 1 for training, with other five samples from the first seven images of session 2 for testing. In the CMU Multi-PIE dataset, we randomly choose one–five samples per subject as the training set and 10 samples as the test set. The results of our MASR-SRW, MASR-voting, and MASR-NoP methods, as well as other compared approaches in the three datasets are tabulated in Tables II–IV, respectively. Note that since the Volterrafaces method requires at least two training samples per subject for the kernel training, we do not report its results for only one training sample condition. As can be observed, the proposed MASR-SRW method generally achieves the highest recognition rate on all experiments with the training sample size increasing from 1 to 5. It is worth noting that the MASR-SRW, MASR-voting, MJSR-voting, and MJSR-SRW methods perform better than the GSRC-p and SRC-p approaches, indicating the advantage of jointly utilizing the Gabor scale and orientation information. In addition, compared with the MJSR-voting and MJSR-SRW, MASR-SRW and MASR-voting methods generally deliver better results.

TABLE VI
RECOGNITION ACCURACY (%) ON THE AR DATASET WITH OCCLUSION

| Method | Experimental Setting A | | | | Experimental Setting B | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Occlusion Type | | | | Training number | | | | |
| | Sunglass-S1 | Scarf-S1 | Sunglass-S2 | Scarf-S2 | 1 | 2 | 3 | 4 | 5 |
| LBP | 95.5 | 94.8 | 88.3 | 86.5 | 70.6 | 72.3 | 76.5 | 78.3 | 82.5 |
| LGBPHS | 99.3 | 96.6 | 95.6 | 92.3 | 81.6 | 82.0 | 83.4 | 83.9 | 89.3 |
| SRC-p | 92.1 | 88.4 | 88.1 | 85.3 | 59.8 | 66.8 | 72.3 | 76.0 | 80.0 |
| GSRC-NoP | 87.3 | 85.0 | 75.2 | 83.0 | 64.5 | 67.1 | 68.2 | 68.5 | 71.3 |
| GSRC-p | 94.3 | 90.2 | 85.4 | 88.6 | 78.0 | 79.5 | 83.8 | 86.0 | 86.6 |
| RRC | 99.0 | 93.3 | 89.3 | 76.3 | 56.4 | 58.5 | 61.8 | 65.2 | 70.4 |
| SLF-RKR | **100** | **100** | 93.0 | 97.6 | 81.3 | 84.5 | 88.3 | 89.6 | 91.8 |
| Volterrafaces | 90.3 | 92.1 | 68.9 | 72.2 | — | 46.3 | 51.1 | 49.9 | 61.4 |
| MJSR-NoP | 93.3 | 93.7 | 76.3 | 82.5 | 70.1 | 71.2 | 72.5 | 73.8 | 79.0 |
| MJSR-voting | 99.3 | 99.3 | 95.5 | 97.4 | 83.1 | 86.3 | 90.1 | 91.6 | 92.8 |
| MJSR-SRW | **100** | **100** | 96.3 | 98.0 | 85.3 | 88.2 | 91.3 | 92.5 | 94.0 |
| MASR-NoP | 93.6 | 94.2 | 77.2 | 83.7 | 72.3 | 74.3 | 75.3 | 75.5 | 76.6 |
| MASR-voting | **100** | 99.6 | 96.6 | 97.6 | 85.1 | 90.0 | 91.6 | 92.2 | 93.4 |
| MASR-SRW | **100** | **100** | **97.0** | **98.2** | **86.9** | **91.5** | **94.2** | **94.0** | **96.2** |


Fig. 7. Test examples with occlusion and pixel corruption. (a) Extended Yale B test images with 50% simulated occlusion. (b) AR test images with real occlusion (e.g., sunglass and scarf). (c) AR test images with real occlusion and 50% simulated pixel corruption.

Furthermore, we can see that the advantages of the MASR-NoP method over the MJSR-NoP and GSRC-NoP methods become more obvious, which can further validate the effectiveness of adaptive atom selection strategy.

### *D. Face Recognition With Occlusion*

In this section, we test the robustness of the MASR-SRJW, MASR-voting, MASR-NoP, and other compared approaches to occlusion using the Extended Yale B and AR datasets. In the Extended Yale B dataset, experiments of two different settings (A and B) are conducted. The experimental setting A is the same as that in [31] and [46], which chooses subsets 1 and 2 (about 19 samples) for training and subset 3 for testing per subject. In the experiment setting B, a more challenging experimental setting is used, in which three samples are used as the training set and 10 samples as the testing set for each subject. Since samples in the Extended Yale B do not contain occlusion, as in [29], we simulated various levels of occlusion (from 10% to 50%) by replacing a randomly located block in each test image with an unrelated image. Fig. 7(a) shows test examples with 50% occlusion. As mentioned above, each session of the AR dataset includes six images with real occlusions [see examples in Fig. 7(b)]. In the AR dataset, we also used two different experimental settings (A and B). In the setting A, as in [31] and [46], seven nonoccluded samples are used for training, while the remaining samples from both sessions 1 and 2 are chosen for testing. In the setting B, we choose one–five nonoccluded samples from session 1 for training and six occluded samples from session 2 for testing. Note that samples in session 2 are usually harder to be recognized than samples from session 1, since more serious occlusions are existed in the samples of session 2. The recognition rates on these two datasets are reported in Tables V and VI. As can be observed, MASR-SRW and MASR-voting methods still show much better results than the compared approaches. In addition, different from the nonocclusion condition, the MASR-SRW leads to more improvement over the MASR-voting. This is because the proposed SRW strategy can adaptively assign low weights for occlusion regions and thus reduce the occlusion interferences, which verifies the effectiveness of the SRW. Furthermore, though the SLF-RKR and RRC can perform very well under the same experimental setting A of the two datasets, their recognition results are worse than those of the proposed MASR-SRW method under the experimental setting B when smaller number of training samples and more serious occlusion are used. This can further demonstrate the effectiveness of multitask sparse representation and adaptive weighting schemes adopted in the proposed method for these extreme conditions.

### *E. Face Recognition With Both Occlusion and Corruption*

In this section, we consider an extreme condition that further adds pixel corruptions to the occluded AR test samples. First, we randomly choose three nonoccluded samples from session 1 as training set and six occluded samples from session 2 as testing set. Then, for each testing image, we replace a certain percentage of its pixels by uniformly distributed random values within [0, 255]. Fig. 7(c) shows test examples with occlusion and 50% pixel corruption. As can be seen, even for humans, the corrupted face images are difficult to recognize. Table VII lists the results of the MASR-SRW, MASR-voting, MASR-NoP, and other compared methods under occlusions and different percentage of corrupted pixels. We can observe that the MASR-SRW still performs better than the MASR-voting and other compared approaches. For the pixel corruption level from 10% to 30%, the recognition rates of the MASR-SRW are almost 20% higher than those

TABLE VII
RECOGNITION ACCURACY (%) ON THE AR DATASET WITH REAL OCCLUSION AND SIMULATED PIXEL CORRUPTION (FROM 10% TO 50%)

| Method | Pixel corruption (%) | | | | |
|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 |
| LBP | 24.5 | 14.2 | 9.6 | 7.3 | 4.5 |
| LGBPHS | 55.2 | 38.5 | 23.5 | 14.8 | 8.5 |
| SRC-p | 56.5 | 40.5 | 27.1 | 14.8 | 8.5 |
| GSRC-NoP | 58.8 | 42.8 | 33.0 | 24.6 | 17.8 |
| GSRC-p | 61.8 | 47.1 | 34.0 | 20.3 | 13.8 |
| RRC | 51.2 | 47.0 | 35.8 | 26.3 | 16.6 |
| SLF-RKR | 38.8 | 23.3 | 13.6 | 9.7 | 5.2 |
| Volterrafaces | 43.9 | 31.8 | 18.3 | 12.0 | 7.4 |
| MJSR-NoP | 64.6 | 48.3 | 40.0 | 27.3 | 19.8 |
| MJSR-voting | 76.0 | 64.6 | 47.2 | 31.1 | 20.8 |
| MJSR-SRW | 77.8 | 66.1 | 48.9 | 33.2 | 23.3 |
| MASR-NoP | 66.4 | 51.8 | 42.4 | 30.1 | 21.6 |
| MASR-voting | 78.1 | 66.3 | 49.5 | 33.6 | 22.8 |
| MASR-SRW | **80.5** | **69.1** | **53.5** | **36.7** | **25.2** |

of LBP, LGBPHS, SRC-p, GSRC-p, RRC, SLF-RKR, and Volterrafaces. We should note that the LBP, LGBPHS, and SLF-RKR methods do not perform well in these pixel corruption conditions. This is because the LBP, LGBPHS, and SLF-RKR methods all adopt the histogram as the final feature for classification and the histogram is easily interfered by the pixel corruptions. In addition, since the LGBPHS method adopts the Gabor filtering that can reduce the pixel corruptions to some extent, such a method can deliver better results than the LBP and SLK-RKR methods.

## V. CONCLUSION

In this paper, we presented a novel multitask sparse representation method called MASR and a weighting strategy called SRW for Gabor local-feature-based face recognition. The proposed MASR method does not only exploit the intercorrelation among the Gabor features of different scales and orientations but also allows each feature to achieve the better representation via an adaptive atom selection scheme. In addition, by considering the structural characteristics of the face and the effects of the external interferences, the SRW strategy can adaptively combine the decision of each local Gabor region. Extensive experiments demonstrated the effectiveness of the proposed MASR-SRW method on different conditions, including face occlusion, pixel corruption, small number of training samples, as well as variations of lighting and expression.

In addition to the Yale B, AR, and Multi-PIE datasets, we have also tested the proposed MASR-SRW method on a more challenging dataset labeled faces in the wild (LFW). Our obtained recognition accuracy is inferior to that of the very recent DL work [24]. The main reason is that the proposed method cannot handle the unconstrained pose face recognition well. Since the holistic kind of features (e.g., Fourier transform or discrete cosine transform) are more robust to large pose changes, part of our ongoing work is to adaptively fuse the holistic feature with our local Gabor features for the unconstrained pose face recognition problem. In addition, our future work will design a learning algorithm to train a representative and discriminative dictionary from a large number of face samples outside the LFW data. Furthermore, since the proposed method can work very effectively on the small number of training samples condition, in the future, there is strong incentive to apply our MASR model to other recognition applications with very limited reference training data (e.g., pathological structures recognition in medical images and land cover recognition in remote sensing images).

## ACKNOWLEDGMENT

The authors would like to thank the editors and all of the anonymous reviewers for their constructive suggestions that greatly improved this paper.

## REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.
[2] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 947–954.
[3] G. Betta, D. Capriglione, M. Corvino, C. Liguori, and A. Paolillo, "Face based recognition algorithms: A first step toward a metrological characterization," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 1008–1016, May 2013.
[4] H. Sellahewa and S. A. Jassim, "Image-quality-based adaptive face recognition," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 805–813, Apr. 2010.
[5] P. N. Belhumeur, J. P. Hespanha, and D. 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.
[6] P. Comon, "Independent component analysis: A new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
[7] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 166–178, Jan. 2006.
[8] 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.
[9] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 469–481.
[10] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.
[11] C. A. R. Behaine and J. Scharcanski, "Enhancing the performance of active shape models in face recognition applications," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2330–2333, Aug. 2012.
[12] Z. Xu, H. R. Wu, X. Yu, K. Horadam, and B. Qiu, "Robust shape-feature-vector-based face recognition system," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 12, pp. 3781–3791, Dec. 2011.
[13] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 786–791.
[14] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1896, Aug. 2009.
[15] Z. Lei, S. Liao, M. Pietikainen, and S. Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 247–256, Jan. 2011.
[16] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.
[17] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image Vis. Comput.*, vol. 25, no. 5, pp. 553–563, May 2007.
[18] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, Oct. 1980.

[19] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.

[20] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA, USA: Academic, 1990.

[21] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Disc.*, vol. 2, no. 2, pp. 121–167, 1998.

[22] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[23] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2518–2525.

[24] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[25] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996.

[26] M. Shen, Q. Zhang, D. Li, J. Yang, and B. Li, "Adaptive sparse representation beamformer for high-frame-rate ultrasound imaging instrument," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 5, pp. 1323–1333, May 2012.

[27] F. Soldovieri, R. Solimene, L. Lo Monte, M. Bavusi, and A. Loperte, "Sparse reconstruction from GPR data with applications to rebar detection," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 1070–1079, Mar. 2011.

[28] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 927–942, May 2012.

[29] 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.

[30] S. Gao, I. W. Tsang, and L.-T. Chia, "Sparse representation with kernels," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 423–434, Feb. 2013.

[31] M. Yang, L. Zhang, S. C.-K. Shiu, and D. Zhang, "Robust kernel representation with statistical local features for face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 900–912, Jun. 2013.

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

[33] H. Liu, M. Palatucci, and J. Zhang, "Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 649–656.

[34] G. Obozinski, B. Taskar, and M. I. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Statist. Comput.*, vol. 20, no. 2, pp. 231–252, Apr. 2010.

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

[36] X. Mo, V. Monga, R. Bala, and Z. Fan, "Adaptive sparse representations for video anomaly detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 631–645, Apr. 2014.

[37] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Joint dynamic sparse representation for multi-view face recognition," *Pattern Recognit.*, vol. 45, no. 4, pp. 1290–1298, Apr. 2012.

[38] X.-H. Han, X. Qiao, and Y.-W. Chen, "Group sparse representation of adaptive sub-domain selection for image classification," in *Proc. 21st Int. Conf. Pattern Recognit.*, 2012, pp. 1431–1434.

[39] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[40] R. Kumar, A. Banerjee, B. C. Vemuri, and H. Pfister, "Maximizing all margins: Pushing face recognition with kernel plurality," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2375–2382.

[41] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[42] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[43] S. Li and L. Fang, "Signal denoising with random refined orthogonal matching pursuit," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 1, pp. 26–34, Jan. 2012.

[44] L. Fang *et al.*, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2034–2049, Nov. 2013.

[45] B. Efron, T. Hastie, I. Johnston, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[46] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[47] S. Li, L. Fang, and H. Yin, "An efficient dictionary learning algorithm and its application to 3-D medical image denoising," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 417–427, Feb. 2012.

[48] R. Kumar, A. Banerjee, and B. C. Vemuri, "Volterrafaces: Discriminant analysis using Volterra kernels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 150–155.

[49] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Statistics), vol. 1. New York, NY, USA: Springer-Verlag, 2001.

[50] A. Martínez and R. Benavente, "The AR face database," Centre de Visió per Computador, Univ. Autònoma de Barcelona, Barcelona, Spain, Tech. Rep. 24, 1998.

[51] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[52] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, May 2010.

[53] D. Cai, X. He, and J. Han, "Spectral regression: A unified approach for sparse subspace learning," in *Proc. 7th IEEE Int. Conf. Data Mining*, Oct. 2007, pp. 73–82.

**Leyuan Fang** (S'10–M'14) received the B.S. degree in electrical engineering from the Hunan University of Science and Technology, Xiangtan, China, in 2008. He is currently pursuing the Ph.D. degree with the College of Electrical and Information Engineering, Hunan University, Changsha, China.

He has been a Visiting Ph.D. Student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council, since 2011. His current research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Mr. Fang was a recipient of the Scholarship Award for Excellent Doctoral Student by the Chinese Ministry of Education in 2011.

**Shutao Li** (M'07–SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

He joined the College of Electrical and Information Engineering, Hunan University, in 2001. He was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, in 2001. From 2002 to 2003, he was a Post-Doctoral Fellow with Royal Holloway College, University of London, London, U.K., with Prof. J. Shawe-Taylor. In 2005, he visited the Department of Computer Science, Hong Kong University of Science and Technology, as a Visiting Professor. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or co-authored over 160 refereed papers. His current research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li was a recipient of two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and a member of the Editorial Board of the *Information Fusion* journal and the *Sensing and Imaging* journal.