Pattern Recognition ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Joint dynamic sparse representation for multi-view face recognition

Haichao Zhang [a,b,*], Nasser M. Nasrabadi [c], Yanning Zhang [a], Thomas S. Huang [b]

[a] School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
[b] Beckman Institute, University of Illinois at Urbana-Champaign, IL, USA
[c] U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD, USA

ARTICLE INFO

*Article history:*
Received 27 June 2011
Received in revised form
7 September 2011
Accepted 9 September 2011

*Keywords:*
Multi-view face recognition
Joint dynamic sparsity
Joint dynamic sparse representation based classification

ABSTRACT

We consider the problem of automatically recognizing a human face from its multi-view images with unconstrained poses. We formulate the multi-view face recognition task as a joint sparse representation model and take advantage of the correlations among the multiple views for face recognition using a novel joint dynamic sparsity prior. The proposed joint dynamic sparsity prior promotes shared joint sparsity patterns among the multiple sparse representation vectors at *class-level*, while allowing distinct sparsity patterns at *atom-level* within each class to facilitate a flexible representation. Extensive experiments on the CMU Multi-PIE face database are conducted to verify the efficacy of the proposed method.



## 1. Introduction

Person identification is of paramount importance in many applications, such as surveillance and security. Due to the rich information contained in face images, face recognition is still one of the most important approaches for person identification. Much work on face recognition has focused on using a single probe face image for identification [1,2]. These methods are limited by their sensitivity to pose variation of the probe face image, which is a common problem in real world scenarios. Moreover, single image-based recognition methods are not reliable if the input image is noisy or occluded, as typically encountered in practice. In many cases, multiple views of the same face can be obtained for recognition, as in video surveillance, where multiple snapshots at different time instances capturing varying poses of the subject are available. Similarly, in video camera networks, multiple images of the same subject at different viewpoints are available for identification. Therefore, it is natural to exploit the paradigm of using multiple face images for recognition. Currently, most of the existing face recognition techniques are designed for single frontal view based face recognition, which are obviously not optimal in the multi-view scenario due to their sensitivity to poses and the failure of exploiting the inter-correlation among the multiple views of the same subject. In this paper, we will develop a novel face recognition method that can take multiple face images captured from possibly different viewpoints for recognition, without requiring the poses/viewpoints to be known or estimated.

Various kinds of algorithms have been developed for face recognition in the literature. Nearest Neighbor (NN) method is one of the most simple and intuitive methods for face recognition. NN classifies the probe face based on the best representation using a single training sample, which is essentially a point-to-point classification method. The Nearest Subspace (NS) method [3,4] generalizes NN method in the sense that it classifies the test sample based on the best linear representation in terms of all the training samples in each class. In this way, the classification decision can be made by using the information from all the training samples of each class, which is more robust than NN. The sparse representation-based classification (SRC) method [5] is a further generalization of NS by representing the test sample using the training samples (atoms) adaptively selected from all the candidate training samples (a structured dictionary) from both within and across different classes. This method has been demonstrated to be more robust to illumination and sparse corruption, which are the common factors that degrade the performance of the face recognition algorithms. By nature, NN, NS, and SRC are all *single* image-based face recognition methods, which only use the information from a single input face for recognition. For some recent advances in single image-based face recognition, refer to [6–8].

Recently, there has been a growing interest in face recognition from a set of images due to its advantages over single image-based methods. By using multiple face images of the same subject for recognition, it can potentially improve the robustness of the recognition system to different kinds of variations. Several different schemes have recently been developed in the literature. In [9],

* Corresponding author.
*E-mail address:* hczhang1@gmail.com (H. Zhang).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.09.009

a single view image obtained via the weighted average of the multi-view face images is used for achieving multi-view recognition. Inspired by label propagation [10], a graph-based face recognition method using image set is proposed in [11], which first constructs a similarity graph between the images in the input set and those in the training set, and then applies a class-wise graph matching procedure for joint identification. Another set-based face recognition method is proposed in [12], where each face set is modeled by affine hulls (or convex hulls) spanned by the samples in the set, and the classification decision is based on the distance from the affine hull of observations to that of training samples from each class. Among all the approaches, one of the most well-known set-based face recognition approaches is the Mutual Subspace Method (MSM) [13], which models the face set as a point on a Grassmann manifold [14], thus converting the task of comparing set-to-set distance to point-to-point distance on the Grassmann manifold, where distance-based classifiers can be applied. The above-mentioned methods are all limited by the fact that they calculate the distance between the input test set and each of the gallery face sets independently. Moreover, they treat each set of gallery face images as a single linear subspace, which may not be proper when large pose variations exist. A natural generalization of the SRC method is the joint sparse representation-based classification (JSRC) method which can be applied for classification in the presence of multiple test samples [15,16]. The JSRC method assumes that the multiple test samples share the same sparsity pattern. However, this is inappropriate for multi-view faces with large pose differences, thus limiting its practical applications.

In this paper, we propose a novel Joint Dynamic Sparse Representation based Classification (JDSRC) method for multi-view face recognition. A graphical illustration of the proposed method is depicted in Fig. 1. The proposed method exploits the correlations (or relationships) among the multiple views using a novel concept of *joint dynamic sparsity*, thus improving the overall performance of a recognition system without requiring any post-processing. Moreover, the proposed method allows a more flexible atom selection process than the JSRC method [15,16] and also has the advantage of not requiring the poses to be known or estimated. As shown in Fig. 1, given a set of observations from different viewpoints for the same subject, we first perform a joint dynamic sparse representation of this observation ensemble with respect to a dictionary of training images and then classify the observation ensemble to the class which gives the minimum total reconstruction error. As the multiple observations describe the same subject, the recovered sparse representation vectors tend to have the same sparsity pattern at class-level, ideally with non-zero coefficients only associated with the training images belonging to the correct subject (class) in the dictionary. Furthermore, since the multiple view face images are captured from quite different viewpoints, the atom-level sparsity patterns of the representation vectors are not going to be the same for all the multi-view probe faces, thus the non-zero coefficients tend to be mainly associated with the training images of similar viewpoints to each probe face that belong to the same class. We term this property as the *joint dynamic sparsity* when multiple sparse representation vectors share sparsity patterns at class-level but not necessarily at atom-level. Respecting this property, the proposed JDSRC method can achieve several significant goals: (1) it exploits the correlations among all the views and combines the information from each view for discrimination during the joint sparse recovery process rather than performing a post-processing procedure, thus potentially avoiding the risk of making erroneous decisions for each observation when treated independently; (2) the joint dynamic sparsity model adopted in JDSRC is also more flexible and adaptive than the atom selection procedure used in simultaneous/joint sparse representation [15,16], thus is more effective for multi-view face recognition task.

This paper is an extension of our previous work reported in [17], where we have developed a sparsity-based classification algorithm that was applied to multiple measurements obtained from a set of heterogeneous sensors. In that paper we addressed the issues on the general classification problems which can deal with heterogeneous sensors and did not address the specific issues on face recognition problems thoroughly. However, in this paper we focus specifically on the multi-view face recognition problem where we have conducted extensive experiments to evaluate the performance of our proposed algorithm on a large multi-view face database, consisting of a large number of subjects as well as involving multiple views from a set of homogenous sensors. Furthermore, we examine the effects of sparsity, recognition performance in the presence of large pose differences as well as time complexity of the algorithm, which are not evaluated in [17]. The results are compared with several state-of-the-art face recognition methods in the literature.

The rest of the paper is organized as follows. In Section 2, we make a brief review of sparse representation-based method for single view-based face recognition. Then we derive the novel joint dynamic sparse representation-based multi-view face recognition method in Section 3, accompanied with a detailed description of the implementation of the proposed method. Extensive experiments are conducted using the CMU Multi-PIE database [18] in Section 4 under various settings to evaluate the efficacy of the proposed method compared with both classical as well as *state-of-the-art* methods. Finally, we make some discussions and conclude this paper in Section 5.

## 2. Face recognition via sparse representation

The task of recovering the sparse representation of a datum with respect to a basis or a dictionary has been an active topic in many fields, such as image processing [19], computer vision, and pattern recognition [20]. Sparsity is the key factor of recent active

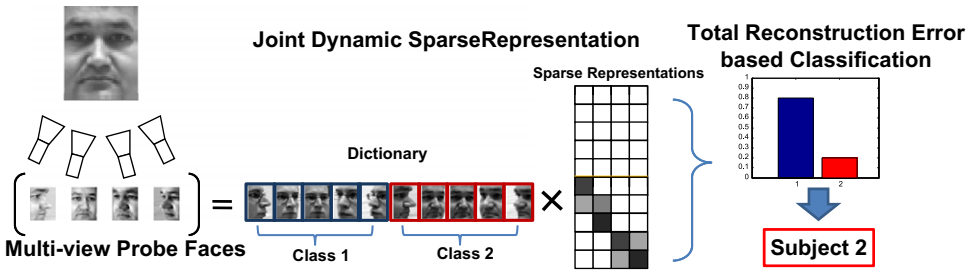


**Fig. 1.** Multi-view face recognition framework. Each subject is imaged from different viewpoints, generating multi-view probe faces, which is the input to our face recognition system. This multi-view probe face set is then used for joint dynamic sparse representation. Finally, the class-wise reconstruction errors are calculated, and the class with the minimum reconstruction error is regarded as the label for the probe subject.

Please cite this article as: H. Zhang, et al., Joint dynamic sparse representation for multi-view face recognition, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.09.009

topics such as sparse coding [21,22] and compressive sensing [23,24], which has conventionally been used as a strong prior for alleviating the ill-posedness of the inverse problems [19]. But recent work shows that the sparse coefficients are also discriminative [5,22]. Recently, a sparse representation based classification method for a single image frontal view-based face recognition is developed in [5]. This method casts the task of face recognition as one of classifying between several linear regression face models via sparse representation, and is based on a simple assumption that data samples of the same class lie in the same subspace of a much lower dimensionality. Thus, a new test sample, $\mathbf{y}$, from class $i$ lies in the same subspace as the training samples (atoms) of the same class $\mathbf{A}^i=[\boldsymbol{\alpha}_1^i,\boldsymbol{\alpha}_2^i,\ldots]$; therefore, it can be represented as a linear combination of the samples from class $i$ as

$$\mathbf{y}=x_1^i\boldsymbol{\alpha}_1^i+x_2^i\boldsymbol{\alpha}_2^i+\cdots=\mathbf{A}^i\mathbf{x}^i, \tag{1}$$

where $\mathbf{x}^i=[x_1^i,x_2^i,\ldots]$ is the coefficient vector containing the appropriate weights for each atom in class $i$. This naturally leads to a sparse representation over the whole training dataset of all the $C$ classes:

$$\mathbf{y}=\mathbf{A}\mathbf{x}=[\mathbf{A}^1\ \mathbf{A}^2\ \cdots\ \mathbf{A}^i\ \cdots\ \mathbf{A}^C]\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \vdots\\ \mathbf{x}^i\\ \vdots\\ \mathbf{0}\end{bmatrix}, \tag{2}$$

where $\mathbf{A}=[\mathbf{A}^1\ \mathbf{A}^2\ \cdots\ \mathbf{A}^i\ \cdots\ \mathbf{A}^C]$ is a structured dictionary consisting of $C$ class-subdictionaries, and $\mathbf{x}=[\mathbf{0}^\top\ \mathbf{0}^\top\ \cdots\ \mathbf{x}^{i\top}\ \cdots\ \mathbf{0}^\top]^\top$ is the corresponding sparse representation vector. As the class label for the test image $\mathbf{y}$ is unknown, it is assumed that its representation, $\mathbf{x}$, with respect to the whole training set, $\mathbf{A}$, can be recovered via the sparse representation procedure shown in [5]:

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

or via the sparsity constrained form

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$$
$$\text{s.t. } \|\mathbf{x}\|_0\le K, \tag{4}$$

where $\varepsilon$ is the reconstruction error parameter and $K$ is the sparsity level, representing the number of active atoms in the dictionary (*i.e.*, those atoms with non-zero coefficients). The class label is decided based on the minimum reconstruction error criteria as

$$\hat{i}=\text{SRC}(\mathbf{y})=\arg\min_i\|\mathbf{y}-\mathbf{A}\delta^i(\hat{\mathbf{x}})\|_2=\arg\min_i\|\mathbf{y}-\mathbf{A}^i\hat{\mathbf{x}}^i\|_2, \tag{5}$$

where $\delta^i(\hat{\mathbf{x}})$ is an operator that keeps the elements of $\hat{\mathbf{x}}$ corresponding to the $i$-th class, while setting others to be zero. This SRC method has been shown to achieve superior performance on single frontal view based face recognition [5].

## 3. Multi-view face recognition via joint dynamic sparse representation

We will present our novel multi-view face recognition method in this section. Given $N^i$ training images from class $i$, which may be captured from different viewpoints, denoted as a class-subdictionary $\mathbf{A}^i=[\boldsymbol{\alpha}_1^i,\ldots,\boldsymbol{\alpha}_{N^i}^i]\in\mathbb{R}^{d\times N^i}$, we can then collect the class-subdictionaries for all the $C$ classes into a single structured dictionary as: $\mathbf{A}=[\mathbf{A}^1,\mathbf{A}^2,\ldots,\mathbf{A}^C]\in\mathbb{R}^{d\times N}$, where $d$ is the dimensionality of data and $N=\sum_{i=1}^C N^i$ is the total number of training samples. Given $M$ views $\{\mathbf{y}_1,\ldots,\mathbf{y}_M\}$ from different viewpoints of a test subject, we arrange them column-wise into a view matrix as $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_M]\in\mathbb{R}^{d\times M}$ and denote their sparse representations with respect to $\mathbf{A}$ as a coefficient matrix $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_M]\in\mathbb{R}^{N\times M}$.

### 3.1. Joint sparse representation model revisited

In many practical situations, we have access to multiple views of the same subject. Therefore, it is natural to exploit the information from all the multiple views to make a single joint decision for recognition task. Given $M$ views of the same subject, we can rewrite all the $M$ sparse representation problems (4) together as

$$\{\hat{\mathbf{x}}_i\}_{i=1}^M=\arg\min_{\{\mathbf{x}_i\}}\sum_{i=1}^M\|\mathbf{y}_i-\mathbf{A}\mathbf{x}_i\|_2^2$$
$$\text{s.t. } \|\mathbf{x}_i\|_0\le K,\quad\forall 1\le i\le M. \tag{6}$$

However, this formulation does not exploit the relationship between the different views since the minimization is separable over each view (Fig. 2(a)). To combine the information from multiple views for recognition, a joint sparsity assumption (constraint) can be applied to the representation vectors [16], which states that the multiple sparse representation vectors share the same sparsity pattern, *i.e.*, selecting the same set of atoms for representation for each view, while the coefficient values corresponding to the same selected atoms may be different, as shown in Fig. 2(b). Under this assumption, the sparse representations for the multiple views are recovered jointly by solving the following optimization problem:

$$\hat{\mathbf{X}}=\arg\min_{\mathbf{X}}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_{\mathcal{F}}^2$$
$$\text{s.t. } \|\mathbf{X}\|_{\ell_0\backslash\ell_2}\le K, \tag{7}$$

where $\|\cdot\|_{\mathcal{F}}^2$ denotes the Frobenius norm. $\|\mathbf{X}\|_{\ell_0\backslash\ell_2}$ is the mixed-norm by applying $\ell_2$-norm on each row of $\mathbf{X}$ and then applying $\ell_0$-norm on the resulting vector. The classification method with this type of joint sparsity prior is called the joint sparse representation-based classification method. The sparse representation matrix recovered via (7) will have the property of row sparsity, as illustrated in Fig. 2(b). However, in the case of multi-view face recognition, assuming that all the views share the same sparsity pattern is too restrictive since the multiple views could have been captured from quite different view points. Therefore, all the views cannot be properly represented by the same set of atoms. This is addressed by the novel model presented in the sequel.

### 3.2. Joint dynamic sparse representation model

In reality, it is often the case that due to the variation of observation conditions (*e.g.*, viewpoint, illumination), each view can be better represented by a *different* set of samples but from

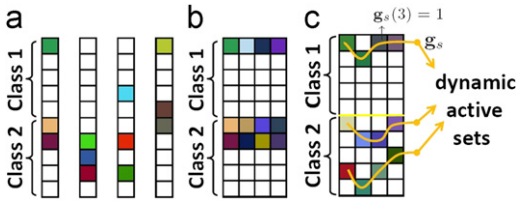


**Fig. 2.** Pictorial illustration of different sparsity models for coefficient matrix $\mathbf{X}$. Each column denotes a sparse representation vector and each squared block denotes a coefficient value. A white block denotes zero entry value. Colored blocks denote non-zeros values. (a) Separate sparsity: each sparse representation vector is treated separately (see Eq. (6)), (b) joint sparsity: the sparse representation vectors share a same sparsity pattern (see Eq. (7)), (c) joint dynamic sparsity: the multiple sparse representation vector share the same sparsity pattern at class-level but atom-level sparsity pattern could be different (see Eq. (10)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: H. Zhang, et al., Joint dynamic sparse representation for multi-view face recognition, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.09.009

the *same* class-subdictionary, *i.e.*, sharing the same sparsity pattern at class-level, but not necessarily at atom-level. Therefore, the desired sparse representation vectors for the multiple observations should share the same class-level sparsity pattern, while their atom-level sparsity patterns may be distinct, *i.e.*, following joint dynamic sparsity, as illustrated in Fig. 2(c). In the following, we will introduce our Joint Dynamic Sparse Representation (JDSR) model, which exploits the joint dynamic sparsity prior for multi-view face recognition. One of the key ingredients in our JDSR model for promoting joint dynamic sparsity is the introduction of the *dynamic active set*. Each dynamic active set, $\mathbf{g}_s \in \mathbb{R}^M$ for $s=1,2,\ldots$, refers to the (row-) indices of a set of coefficients belonging to the same class in the coefficient matrix $\mathbf{X}$, where a number of dynamic active sets are jointly activated during the sparse representation of multiple observations. Each dynamic active set $\mathbf{g}_s$ contains only one index for each column of $\mathbf{X}$, *e.g.*, $\mathbf{g}_s(m)$ is the row-index of the selected atom for the $m$-th column of $\mathbf{X}$ in the $s$-th dynamic active set, as shown in Fig. 2(c). Therefore in our algorithm, we allow the sparse representation for each view to be different, but are forced to share the same *class-level* (group) structure, due to the physical constraint that all the views are from the same subject and belong to the same class.

We formulate our JDSR model as a multivariate regression problem with a novel joint dynamic sparsity promoting term, which is derived in the sequel. The following properties are desired in designing such a term: (i) cues from multiple views should be combined during joint sparse representation, thus enhancing the robustness of joint sparse recovery; (ii) sparsity across dynamic active sets should be promoted, thus inducing joint dynamic sparsity pattern over the recovered multiple sparse representation vectors. To combine the strength of all the atoms within a dynamic active set (thus, across all the views), we apply $\ell_2$-norm over each dynamic active set. To promote sparsity to allow a small number of dynamic active sets to be involved during the joint sparsity representation, we apply $\ell_0$-norm across the $\ell_2$-norm of the dynamic active sets. Therefore, we arrive at the following joint dynamic sparsity promoting term:

$$\|\mathbf{X}\|_G = \|[\|\mathbf{x}_{\mathbf{g}_1}\|_2, \|\mathbf{x}_{\mathbf{g}_2}\|_2, \cdots]\|_0, \tag{8}$$

where $\mathbf{x}_{\mathbf{g}_s}$ denotes the vector formed as the collection of the coefficients associated with the $s$-th dynamic active set $\mathbf{g}_s$:

$$\mathbf{x}_{\mathbf{g}_s} = \mathbf{X}(\mathbf{g}_s) = [\mathbf{X}(\mathbf{g}_s(1),1), \mathbf{X}(\mathbf{g}_s(2),2), \cdots, \mathbf{X}(\mathbf{g}_s(M),M)]^\top \in \mathbb{R}^M. \tag{9}$$

To recover the sparse representation coefficient matrix $\mathbf{X}$ with the joint dynamic sparsity constraint for the multiple observations $\{\mathbf{y}_m\}_{m=1}^M$, we propose the following Joint Dynamic Sparse Representation (JDSR) model:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_{\mathcal{F}}^2$$
$$\text{s.t. } \|\mathbf{X}\|_G \le K, \tag{10}$$

where $K$ is the sparsity level. The use of joint dynamic sparsity regularization term $\|\mathbf{X}\|_G$ has the following advantages:

- $\ell_2$-norm is applied over each dynamic active set, thus allowing to combine the cues from all the views during joint sparse representation; moreover, dynamic active sets are very flexible in selecting the atoms of the same class from the dictionary, and they will provide a better representation of the multiple view images, which are the different measurements of the same subject from different view-points;
- $\ell_0$-norm is applied across the dynamic active sets, thus encouraging the selection of the most parsimonious and representative dynamic active sets, which promotes a joint sparsity pattern shared at class-level while allowing the within-class sparsity patterns to be distinct in order to facilitate the class-wise selection of the most representative atoms for each view.

### 3.3. Joint dynamic sparse representation algorithm

The JDSR model (10) is very challenging to solve due to the co-existence of $\ell_0$-norm and joint dynamic sparsity constraint. We propose to solve (10) with a greedy JDSR algorithm, as detailed in Algorithm 1. The proposed JDSR algorithm has a similar algorithmic structure as Simultaneous Orthogonal Matching Pursuit (SOMP) [16] and Compressive Simultaneous Orthogonal Matching Pursuit (CoSOMP) [25], which includes the following general steps: (i) select new candidates based on the current residue; (ii) merge the newly selected candidate set with the previously selected atom set; (iii) estimate the representation coefficients based on the merged atom set; (iv) prune the merged atom set to a specified sparsity level based on the newly estimated representation coefficients; (v) update the residue. This procedure is iterated until certain conditions are satisfied [25]. We use $\mathbf{X}(:,i)$ to denote the $i$-th column of $\mathbf{X}$ and use $\mathbf{X}(:,\mathbf{i})$ to denote all the columns indexed by $\mathbf{i}$ (similar notations are used for the rows). The major difference between our proposed JDSR algorithm and CoSOMP [25] lies in the atom selection criteria used in steps (i) and (iv) of Algorithm 1, which is detailed in the sequel.

**Algorithm 1.** Joint Dynamic Sparse Representation (JDSR).

**Input:** multi-view data matrix $\mathbf{Y}$, dictionary $\mathbf{A}$, sparsity level $K$, number of views $M$.
**Output:** sparse coefficients matrix $\mathbf{X}$.
**Initialize:** $\mathbf{R} \leftarrow \mathbf{Y}$, $\mathbf{I} \leftarrow \emptyset$
**While** stopping criteria false **Do**
  $\mathbf{E} = \mathbf{A}^\top \mathbf{R}$
  % **(i)** atom selection via joint dynamic sparse mapping
  $\mathbf{I}_{\text{new}} \leftarrow \mathbb{P}_{\text{JDS}}(\mathbf{E}, 2K)$
  $\mathbf{I} \leftarrow [\mathbf{I}^\top, \mathbf{I}_{\text{new}}^\top]^\top$ % **(ii)**index matrix updating
  % **(iii)** representation coefficients updating
  **For** $m = 1,2,\ldots,M$
    $\mathbf{i} \leftarrow \mathbf{I}(:,m)$
    $\mathbf{C}(\mathbf{i},m) \leftarrow (\mathbf{A}(:,\mathbf{i})^\top \mathbf{A}(:,\mathbf{i}))^{-1} \mathbf{A}(:,\mathbf{i})^\top \mathbf{Y}(:,m)$
  **End**
  % **(iv)** atom pruning via joint dynamic sparse mapping
  $\mathbf{I} \leftarrow \mathbb{P}_{\text{JDS}}(\mathbf{C}, K)$ % joint dynamic sparse mapping
  $\mathbf{X} \leftarrow \mathbf{0}$
  **For** $m = 1,2,\ldots,M$
    $\mathbf{i} \leftarrow \mathbf{I}(:,m)$, $\mathbf{X}(\mathbf{i},m) \leftarrow \mathbf{C}(\mathbf{i},m)$
  **End**
  $\mathbf{R} = \mathbf{A}\mathbf{X} - \mathbf{Y}$ % **(v)** residue updating
**End While**
$\mathbf{X} \leftarrow \mathbf{0}$
**For** $m = 1,2,\ldots,M$
  $\mathbf{i} \leftarrow \mathbf{I}(:,m)$
  $\mathbf{X}(\mathbf{i},m) \leftarrow (\mathbf{A}(:,\mathbf{i})^\top \mathbf{A}(:,\mathbf{i}))^{-1} \mathbf{A}(:,\mathbf{i})^\top \mathbf{Y}(:,m)$
**End**

At each iteration of JDSR (step (i) and (iv)), given a coefficient matrix $\mathbf{Z} \in \mathbb{R}^{N \times M}$, we need to select $L$ most representative dynamic active sets from $\mathbf{Z}$, *i.e.*, constructing the best approximation $\hat{\mathbf{Z}}_L$ to $\mathbf{Z}$ with $L$ dynamic active sets (*i.e.*, $\|\hat{\mathbf{Z}}_L\|_G = L$). This can be obtained as the solution to the following problem:

$$\hat{\mathbf{Z}}_L = \arg\min_{\mathbf{Z} \in \mathbb{R}^{N \times M}} \|\mathbf{Z}-\mathbf{Z}_L\|_{\mathcal{F}}$$
$$\text{s.t. } \|\mathbf{Z}_L\|_G \le L. \tag{11}$$

Please cite this article as: H. Zhang, et al., Joint dynamic sparse representation for multi-view face recognition, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.09.009

The solution to (11) can be obtained by a procedure called the Joint Dynamic Sparsity mapping (JDS mapping):

$$\mathbf{I}_L = \mathbb{P}_{\mathrm{JDS}}(\mathbf{Z}, L), \tag{12}$$

which gives the index matrix $\mathbf{I}_L \in \mathbb{R}^{L\times M}$ containing the top-$L$ dynamic active sets for all the $M$ views, as detailed in Algorithm 2. For each iteration of the JDS mapping, it will select a new dynamic active set, which is achieved via three steps: (i) find the maximum absolute coefficient for each class and each view; (ii) combine the maximum absolute coefficients across the views for each class as the total response; (iii) select the dynamic active set as the one that gives the maximum total response. After a joint dynamic active set is determined, we keep a record of the selected indices as one row of $\mathbf{I}_L$ and set the associated coefficients in the coefficient matrix to be zero to ensure none of the coefficients will be selected again. This procedure is iterated until the specified number of dynamic active sets are determined. After that, $\hat{\mathbf{Z}}_L$ can be obtained by keeping the entries of $\mathbf{Z}$ selected by $\mathbf{I}_L$ and setting the remaining entries to be zero. As mentioned above, Algorithm 2 is used as a sub-routine for dynamic active set selection for each iteration of Algorithm 1, and this iteration process is repeated on the residue until certain conditions are satisfied [16,25].

**Algorithm 2.** Joint Dynamic Sparsity Mapping $\mathbb{P}_{\mathrm{JDS}}(\mathbf{Z}, L)$.

**Input:** coefficient matrix $\mathbf{Z}$, desired number of dynamic active sets $L$, dictionary atom label vector $\mathbf{u}$, number of classes $C$, number of views $M$.
**Output:** index matrix $\mathbf{I}$ for the top-$L$ dynamic active sets.
**Initialize:** $\mathbf{I} \leftarrow \emptyset$ % initialize the index matrix as empty
**For** $l = 1, 2, \ldots, L$
  **For** $i = 1, 2, \ldots, C$
    $\mathbf{c} \leftarrow \mathrm{find}(\mathbf{u}, i)$ % get the index vector for the $i$-th class
    **For** $m = 1, 2, \ldots, M$
    % **(i)** find the maximum coefficient value $v$ and its index $t$ for the $i$-th class, $m$-th view
      $[v, t] \leftarrow \max(|\mathbf{Z}(\mathbf{c}, m)|)$
      $\mathbf{V}(i, m) \leftarrow v$, $\tilde{\mathbf{I}}(i, m) \leftarrow \mathbf{c}(t)$
    **End**
    % **(ii)** combine the max-coefficients for each class
    $\mathbf{s}(i) \leftarrow \sqrt{\sum_{m=1}^{M} \mathbf{V}(i, m)^2}$
  **End**
  $[\hat{v}, \hat{t}] = \max(\mathbf{s})$ % find the best cluster of atoms belonging to the same class across all the classes
  $\mathbf{I}(l, :) = \tilde{\mathbf{I}}(\hat{t}, :)$, $\mathbf{Z}(\tilde{\mathbf{I}}(\hat{t}, :)) \leftarrow \mathbf{0}^{\top}$
**End**

Assuming the number of training samples for each class in the dictionary is $n$ with the total number of class as $C$ ($N = Cn$), the JDS mapping with $L$-dynamic active sets has the computational complexity of $\mathcal{O}(LMN \log n + LC(M+1))$ for each iteration. Empirical evaluations on the computational complexity have been conducted in Section 4.7.

### 3.4. Recognition criteria

After recovering the sparse representation matrix, $\hat{\mathbf{X}}$, for all the views of the same subject, $\mathbf{Y}$, we make a single decision on the class label for all the views simultaneously based on $\hat{\mathbf{X}}$ by combining the residuals from all the multi-view images:

$$\hat{i} = \arg\min_i \|\mathbf{Y} - \mathbf{A}\delta^i(\hat{\mathbf{X}})\|_{\mathcal{F}}^2. \tag{13}$$

Here $\delta^i(\hat{\mathbf{X}})$ is reused as a matrix operator, keeping the values of $\hat{\mathbf{X}}$ corresponding to the $i$-th class while setting others to be zero. The use of Frobenius norm $\|\cdot\|_{\mathcal{F}}$ combines the reconstruction errors from all the views. The overall procedure of the proposed JDSRC algorithm is summarized in Algorithm 3. When $M = 1$, the proposed JDSRC algorithm reduces naturally to the conventional SRC method and when we confine the indices represented by a dynamic active set to be within the same row of the coefficient matrix, then our algorithm reduces to JSRC method.

**Algorithm 3.** Joint Dynamic Sparse Representation based Classification (JDSRC).

**Input:** multi-view observation set $\mathbf{Y}$, dictionary $\mathbf{A}$, sparsity level $K$
**Output:** class label $\hat{i}$
**Perform** Joint Dynamic Sparse Representation:
$\hat{\mathbf{X}} = \mathrm{JDSR}(\mathbf{A}, \mathbf{Y}, K)$
**Perform** reconstruction: $\hat{\mathbf{Y}}_i = \mathbf{A}\delta^i(\hat{\mathbf{X}})$
**Calculate** residue matrix: $\mathbf{E}_i = \mathbf{Y} - \hat{\mathbf{Y}}_i$
**Infer** class label: $\hat{i} = \arg\min_i \|\mathbf{E}_i\|_{\mathcal{F}}^2$.

## 4. Experiment results

In this section, we carry out extensive experiments under various conditions to evaluate the performance of the proposed JDSRC method. Experiments are conducted on the CMU Multi-PIE database [18], which contains a large number of face images under different illuminations, view points and expressions, up to four sessions over the span of several months. Subjects were imaged under 20 different illumination conditions, using 13 cameras at head height, spaced at 15° intervals. Illustrations for the multiple camera configurations, as well as the captured multi-view images, are shown in Fig. 3. We use the multi-view face images with neutral expressions under varying illuminations for 129 subjects for experiment, which are present in all four sessions. The face regions for all the poses are extracted manually and are resized to $30 \times 23$. The images captured from all the 13 different poses with the view angles $\Theta = \{0°, \pm 15°, \pm 30°, \pm 45°, \pm 60°, \pm 75°, \pm 90°\}$

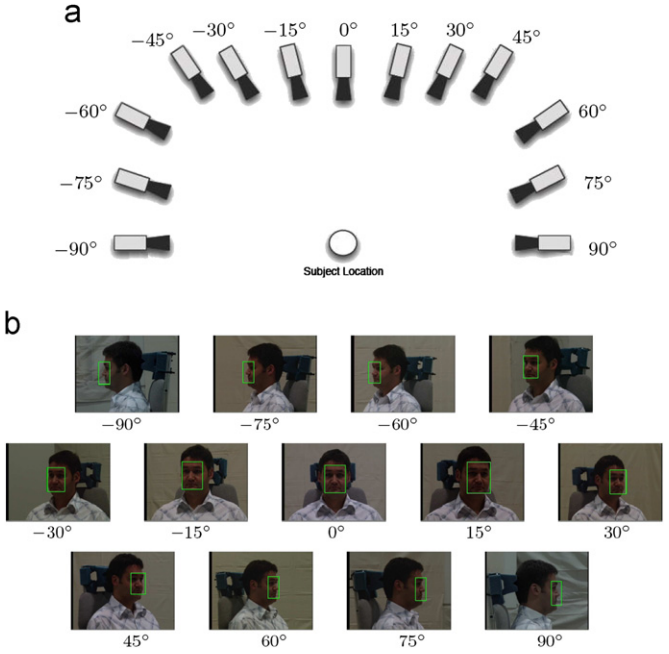


**Fig. 3.** Illustration of the multi-view face images. (a) The configurations of the 13 head-height cameras in Multi-PIE and (b) example multi-view face images captured using the multi-camera configuration shown in (a).

are used for experiment. We compare the proposed method with classical Mutual Subspace Method (MSM) [13], *state-of-the-art* Graph-based method [11], SRC [5] method with majority voting, as well as JSRC method [16]. We set the sparsity level, $K=11$ for SRC, and $K=15$ for JSRC and JDSRC, in all our experiments, and these sparsity levels were empirically found to provide best results, generally. Random projection is used for dimensionality reduction [5] in our experiments.

### 4.1. Face recognition with increasing number of views

In this subsection, we examine the effectiveness of using multiple views for face recognition. We first examine the face recognition performance using a single face image captured from different viewpoints. In this experiment, training images are from Session 1 using view-subset $\Theta_{\text{train}}=\{0^\circ, \pm 30^\circ, \pm 60^\circ, \pm 90^\circ\}$, while testing images are generated from Session 2 from all the views $\Theta_{\text{test}}=\Theta$. This is a more realistic and challenging setting in the sense that the data sets used for training and testing are collected separately and even not all the poses in the testing sets are available for training. To generate a test sample with $M$ views, we first randomly select a subject, $i\in\{1,2,\ldots,129\}$, from the test set and then select $M\in\{1,2,3,4,5,6,7\}$ different views randomly from $\Theta_{\text{test}}$ captured at the same time instance for subject $i$. Two thousand test samples are generated with this scheme for testing. The recognition results on Session 2 of Multi-PIE database with $d=64$ are summarized in Table 1, and the corresponding plots are shown in Fig. 4(a). It is demonstrated that the multi-view based methods ($M>1$) outperform their single-view counterparts ($M=1$) by a large margin, indicating the advantage of using multiple views in face recognition task. This is natural to expect, as face images from different views offer complementary information for recognition. Therefore, by combining the face images from different views properly, we can potentially achieve better recognition performance than that of just using a single view face image. Also, it is noted that the performance of all the algorithms improves as the number of views is increased. However, the JSRC method does not perform well when the number of randomly chosen testing views is large. The reason is that as the number of testing views increases, the view differences between different views could be large, and the assumption used in JSRC that all the views can be represented by the same set of atoms becomes more and more inaccurate. Furthermore, the proposed method outperforms all the other methods under all the different number of views when $M>1$, which indicates that the proposed method is more effective in exploiting the inter-correlation between the multiple views for achieving a joint classification.

**Table 1**
Recognition rate (%) under different number of views ($C=129, d=64$).

| View ($M$) | MSM [13] | Graph [11] | SRC [5] | JSRC [16] | JDSRC |
|---|---|---|---|---|---|
| 1 | 36.5 | 44.5 | **45.0** | **45.0** | **45.0** |
| 3 | 48.9 | 63.4 | 59.5 | 53.6 | **72.0** |
| 5 | 52.5 | 72.0 | 62.2 | 55.0 | **82.3** |
| 7 | 55.9 | 76.5 | 63.3 | 51.4 | **84.5** |

### 4.2. Face recognition under different feature dimensions

In this subsection, we further examine the effects of data (feature) dimensionality $d$ on the recognition rate using Multi-PIE. The test samples are generated using $\Theta_{\text{train}}=\{0^\circ, \pm 30^\circ, \pm 60^\circ, \pm 90^\circ\}$ with $M=5$. Random projection is used for dimensionality reduction, which has been shown to be effective for face recognition in [5]. We vary the data dimensionality in the range of $d\in\{32,64,128,256\}$, and the plots in Fig. 4(b) give the performance of all the algorithms on Session 2 of Multi-PIE dataset. It is shown that MSM does not perform well when the dimensionality of feature is low. However, as the dimensionality of feature increases, its performance increases quickly. JSRC method performs very similar to MSM in this experiment. The Graph-based method [11] performs relatively well under low dimensionality. However, its performance becomes saturated after $d>64$. Similar saturation phenomena can be observed for SRC. The proposed method outperforms all the compared methods by a large margin and performs the best under all the examined dimensionality of features, which implies the effectiveness of the proposed method for multi-view face recognition task.

### 4.3. Face recognition in the presence of view difference between training and testing

In the above experiments, we have used an experiment setup such that *not all* the testing viewpoints were used in the training process, *i.e.*, recognition performance was evaluated in the presence of view differences between training and testing. In this subsection, we further examine the effects of the view differences

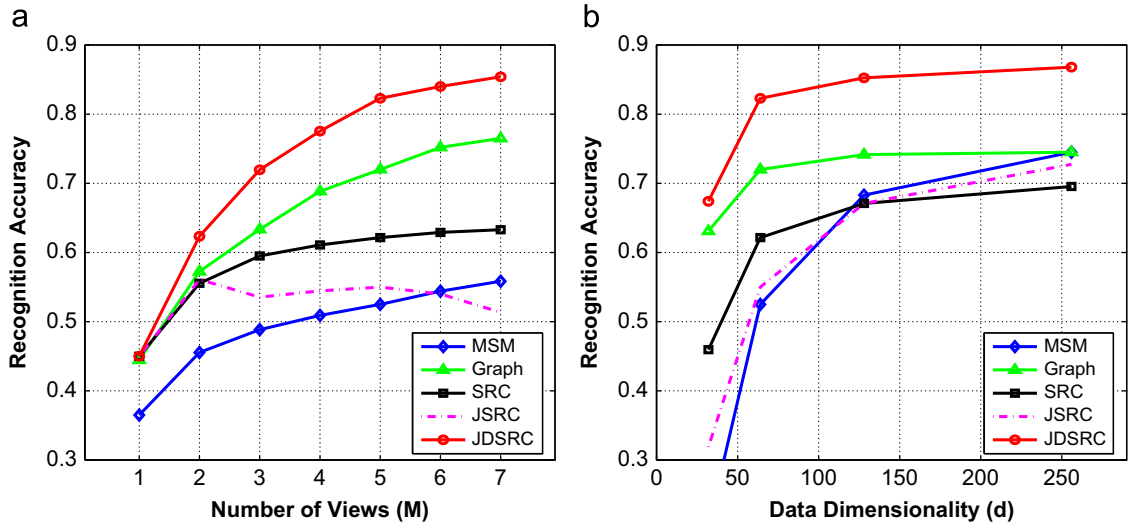


**Fig. 4.** Recognition rate under (a) different number of views with $d=64$ and (b) different feature dimensions with $M=5$.

Please cite this article as: H. Zhang, et al., Joint dynamic sparse representation for multi-view face recognition, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.09.009

on the face recognition performance. Specifically, we examine the recognition performance of having images from a different set of viewpoints in testing where no gallery images are available from those views in the dictionary. We set $M=5$, $d=64$ and use images from Session 1 with the following view angles for training: $\Theta_{\text{train}}=\{0^\circ, \pm 30^\circ, \pm 60^\circ, \pm 90^\circ\}$. Two thousand test samples are generated following the same scheme as described in Section 4.1, but using the following three different view subset selection schemes in order to choose the test subsets: (1) the same views as the training: $\Theta_s=\Theta_{\text{train}}$; (2) completely different views from the training: $\Theta_d=\Theta-\Theta_{\text{train}}$; and (3) mixed view sampling: $\Theta_m=\Theta$ (which is the setup used in Section 4.1). The recognition results of all the algorithms on Session 2 are presented in Table 2. As shown in Table 2, the proposed method performs better under different range of view differences. The Graph-based method [11] performs relatively well under the 'Same views' setting, which is much better than MSM, SRC as well as JSRC. However, as the view difference between training and testing increases (from 'Same views' setting to 'Mixed views' setting and further to 'Different views' setting), Graph-based method [11] degenerates quickly, which implies its sensitivity to the view-differences between training and testing, thus it does not generalize well. The proposed method, on the other hand, is much more robust to the view-differences and outperforms all the other methods significantly, and, thus, is more suitable for real-world applications.

### 4.4. The effects of sparsity

In the sparsity-based recognition method, the sparsity level is an important factor on the recognition performance. In this subsection, we examine the effects of the sparsity level on the recognition performance of the sparsity-based methods. $\Theta_{\text{train}}=\{0^\circ, \pm 30^\circ, \pm 60^\circ, \pm 90^\circ\}$ is used as the training view subset, and $\Theta_{\text{test}}=\Theta_m$ is the test view subset. We vary the sparsity level within the range $K \in \{5,7,9,\ldots,25,27\}$ and examine the recognition rate under each sparsity level for SRC, JSRC and JDSRC. The recognition results with $d=64$ and $M=5$ on Session 2 are shown in Fig. 5. It is noted that for all the recognition methods, the recognition performance first increases with increasing level of sparsity $K$; when the sparsity level surpasses a certain threshold, the recognition performance will be stable or even decrease. The possible reason is that when the sparsity is larger than a certain level, more atoms from the incorrect classes are likely to be selected, thus deteriorating the classification performance. In our experiment, for a specific training view, there are 20 training images per-subject. Therefore, it is reasonable to expect the sparsity level of a test view below this level, which is in accordance with the plots in Fig. 5. This is the reason we set the sparsity level $K=11$ for SRC and $K=15$ for both JSRC and JDSRC, in all the experiments reported in this paper, which give good performances.

### 4.5. Face recognition under large pose differences

In this subsection, we examine more closely the effect of large pose differences on the performance of the multi-view face recognition methods described in this paper. The experimental setups are as follows. We use the face images from all the 13 views in Session 1 for training, *i.e.*, $\Theta_{\text{train}}=\Theta$. For testing, we use $M=3$ views with pose difference from 15° to 90°, *i.e.*, the three testing views are from the pose set $\{0^\circ, \pm\theta\}$, with $\theta \in \{15^\circ,30^\circ,45^\circ,60^\circ, 75^\circ,90^\circ\}$. The dimensionality of data is set as $d=64$. The experimental results are reported graphically in Fig. 6. It can be observed from Fig. 6 that the performance of the MSM method decreases when the pose difference increases. The JSRC method performs better than MSM and it performs well when the pose difference $\theta$ is small, as the assumption that the multiple test views can be represented by the same set of atoms is more appropriate in this scenario. However, when the pose difference is large, the joint sparsity assumption will not hold and becomes inaccurate, thus deteriorating the recognition performance of JSRC. The SRC method with majority voting, on the other hand, appears to be robust with respect to the pose differences and outperforms JSRC when the pose difference is large. The Graph-based method performs better than SRC under different pose differences, at the expense of more computation (see Fig. 7). Our proposed JDSRC method is also not

**Table 2**
Multi-view recognition rate (%) under different view-differences ($C=129$, $d=64$, $M=5$).

| View subset | MSM [13] | Graph [11] | SRC [5] | JSRC [16] | JDSRC |
|---|---|---|---|---|---|
| Same views | 63.5 | 80.7 | 71.1 | 58.6 | **87.5** |
| Mixed views | 52.5 | 72.0 | 62.2 | 55.5 | **82.3** |
| Different views | 35.8 | 46.2 | 47.3 | 36.4 | **66.5** |

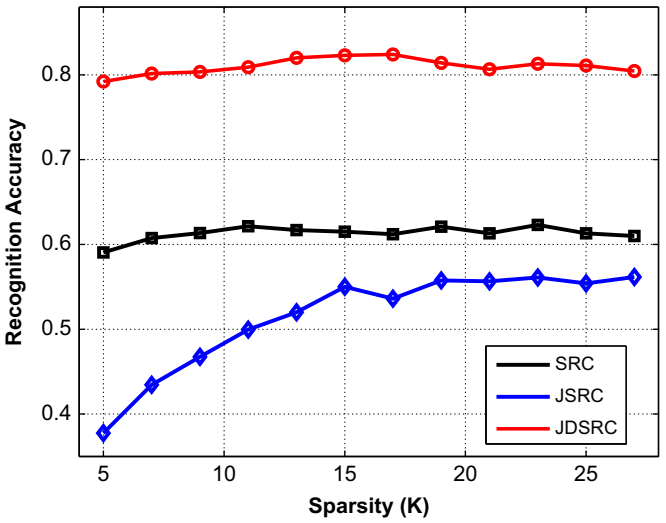


**Fig. 5.** Recognition rate with different sparsity level for SRC, JSRC and JDSRC with $d=64$ and $M=5$.

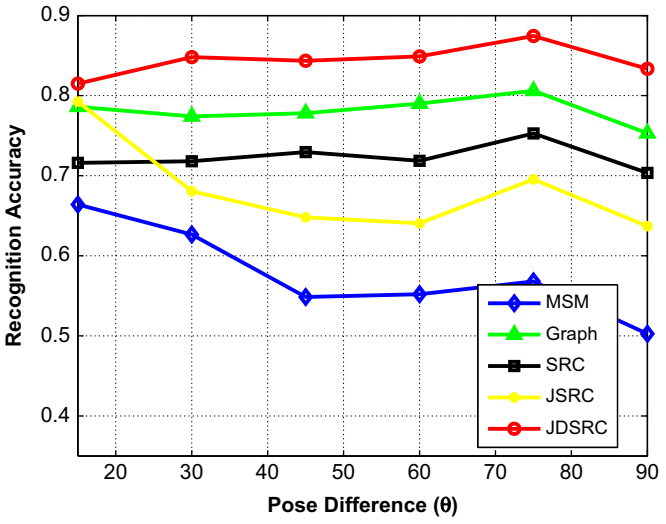


**Fig. 6.** Recognition rate with different pose different for SRC, JSRC and JDSRC with $d=64$ and $M=3$.

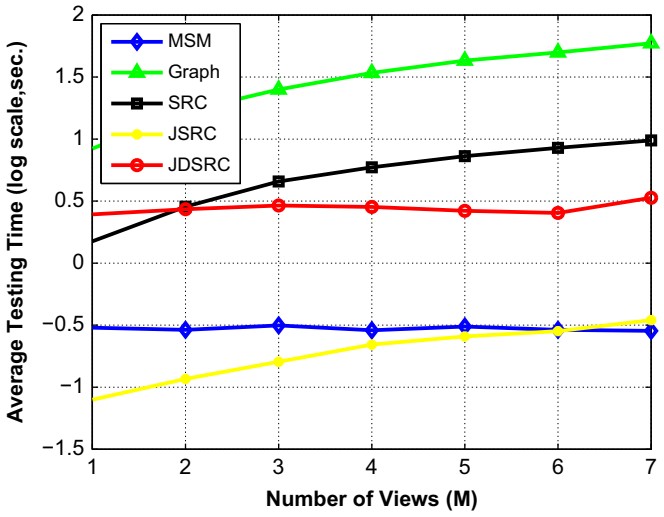


**Fig. 7.** Time complexity comparison for different algorithms (seconds). The time axis is shown in $\log_{10}$ scale.

**Table 3**
Multi-view face recognition rate (%) on different test sessions ($C=129, d=64, M=5$).

| Session | MSM [13] | Graph [11] | SRC [5] | JSRC [16] | JDSRC |
|---|---|---|---|---|---|
| 2 | 52.5 | 72.0 | 62.2 | 55.0 | **82.3** |
| 3 | 49.5 | 65.1 | 56.5 | 48.0 | **77.1** |
| 4 | 45.9 | 62.5 | 52.7 | 45.2 | **73.1** |

sensitive to the pose differences while it can exploit the correlations and the compatible information among the multiple views effectively, outperforming all the other methods, as shown in Fig. 6.

### 4.6. Face recognition under different sessions

In the previous experiments, we used the multi-view face images from Session 1 for training, and the images from Session 2 for testing. To further evaluate the robustness of the proposed method, we present the recognition results using multi-view face images from Session 1 for training and testing on the images captured during several different sessions with the span of several months. Recognition results on Sessions 2–4 data sets with $M=5$ and $d=64$ are summarized in Table 3. It is demonstrated that the proposed method outperforms all the other algorithms under different test sessions. Note that the time interval between Session 1 and other sessions increases with increasing session number, thus increasing the difficulties for face recognition. This explains the gap between the recognition performances of each algorithm on different test sessions.

### 4.7. Time complexity evaluation

Time complexity is an important issue in applications such as face recognition. In this subsection, we perform some empirical evaluation on the time complexity of the proposed method as well as other compared methods. For a fixed number of testing views, we generate 2000 test samples using the approach described in Section 4.1, and record the total time for testing. The average testing time per sample is reported graphically in Fig. 7, where time axis is shown in $\log_{10}$ scale. As shown in Fig. 7, the MSM [13] and JSRC algorithms are the most computationally efficient among all the evaluated algorithms. The Graph-based method [11] is the most computationally intensive one. The algorithm using SRC with majority-voting is also time-consuming when the number of testing views is large. The proposed JDSRC algorithm, although required more computation than MSM, scales much better than the Graph-based method and the SRC method, while achieving the best performance over all the evaluated algorithms (see Figs. 4–6).

## 5. Conclusion

A novel joint dynamic sparse representation-based multi-view face recognition method is presented in this paper. This method inherits the robustness of the sparse representation-based classification method, while also having the advantage of exploiting the inter-correlation among the multiple views. Moreover, the novel joint dynamic sparsity model allows more flexible atom selection for joint sparse representation, which is of vital importance for handling multiple views with different poses for face recognition. Extensive experiments are carried out using the CMU Multi-PIE database. Experimental results of the proposed method were compared with the classical, as well as *state-of-the-art*, methods, and we demonstrated the superior performance of the proposed method on multi-view face recognition task under various variations, including number of views, feature dimensionality, view difference and testing sessions.

## Acknowledgments

We would like to thank all the anonymous reviewers for their constructive comments. This work is supported by the National Natural Science Foundation of China (Nos. 60872145, 60903126), National High Technology Research and Development Program (863) of China (No. 2009AA01Z315), China Postdoctoral (Special) Science Foundation (Nos. 20090451397, 201003685) and Cultivation Fund of the Key Scientific and Technical Innovation Project from Ministry of Education of China (No. 708085). The work is also supported by the U.S. Army Research Laboratory and U.S. Army Research Office under Grant no. W911NF-09-1-0383.

## References

[1] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[3] S.Z. Li, Face recognition based on nearest linear combinations, in: CVPR, 1998, pp. 839–844.
[4] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, D.J. Kriegman, Clustering appearances of objects under varying illumination conditions, in: CVPR, 2003, pp. 11–18.
[5] J. Wright, A.Y. Yang, A. Ganesh, S. Shankar Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.
[6] R. He, W.-S. Zheng, B.-G. Hu, Maximum correntropy criterion for robust face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1561–1576.
[7] R. He, W.-S. Zheng, B.-G. Hu, X.-W. Kong, A regularized correntropy framework for robust pattern recognition, Neural Computation 23 (8) (2011) 2074–2100.
[8] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: CVPR, 2011, pp. 625–632.
[9] M. Farhat, A. Alfalou, H. Hamam, C. Brosseau, Double fusion filtering based multi-view face recognition, Optics Communications 282 (11) (2009) 2136–2142.
[10] D. Zhou, O. Bousquet, T. Navin Lal, J. Weston, B. Schölkopf, Learning with local and global consistency, in: Neural Information Processing Systems, 2004, pp. 321–328.
[11] E. Kokiopoulou, P. Frossard, Graph-based classification of multiple observation sets, Pattern Recognition 43 (12) (2010) 3988–3997.
[12] H. Cevikalp, B. Triggs, Face recognition based on image sets, in: CVPR, 2010, pp. 2567–2573.

[13] K. Fukui, O. Yamaguchi, Face recognition using multi-viewpoint patterns for robot vision, Springer Tracts in Advanced Robotics 15 (2005) 192–201.
[14] J. Hamm, D.D. Lee, Grassmann discriminant analysis: a unifying view on subspace-based learning, in: ICML, 2008, pp. 376–383.
[15] A. Rakotomamonjy, Surveying and Comparing Simultaneous Sparse Approximation (or Group Lasso) Algorithms, Technical Report, 2010.
[16] J.A. Tropp, A.C. Gilbert, M.J. Strauss, Algorithms for simultaneous sparse approximation. Part I: greedy pursuit, EURASIP Journal on Applied Signal Processing 86 (2006) 572–588.
[17] H. Zhang, N.M. Nasrabadi, Y. Zhang, T.S. Huang, Multi-observation visual recognition via joint dynamic sparse representation, in: ICCV, 2011.
[18] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-PIE, Image and Vision Computing 28 (5) (2010) 807–813.
[19] M. Elad, M.A.T. Figueiredo, Y. Ma, On the role of sparse and redundant representations in image processing, Proceedings of the IEEE 98 (6) (2010) 972–982 (Special Issue on Applications of Sparse Representation & Compressive Sensing).
[20] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, Proceedings of the IEEE 98 (6) (2010) 1031–1044 (Special Issue on Applications of Sparse Representation & Compressive Sensing).
[21] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, in: Neural Information Processing Systems, 2006, pp. 801–808.
[22] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: CVPR, 2009, pp. 1794–1801.
[23] E. Candè, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory 52 (2) (2006) 489–509.
[24] D. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006) 1289–1306.
[25] M.F. Duarte, V. Cevher, R.G. Baraniuk, Model-based compressive sensing for signal ensembles, in: Proceedings of the Allerton Conference on Communication, Control, and Computing 2009, pp. 585–600.

**Haichao Zhang** received the B.Eng. degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2007. He is currently pursuing the Ph.D. degree in School of Computer Science, Northwestern Polytechnical University, Xi'an, China. His research interests include the sparse representation and its applications in pattern recognition and signal processing.

**Nasser M. Nasrabadi** received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from the Imperial College of Science and Technology (University of London), London, England, in 1980 and 1984, respectively. From October 1984 to December 1984, he worked with IBM (UK) as a Senior Programmer. During 1985–1986, he worked with Philips Research Laboratory in NY as a Member of the Technical Staff. From 1986 to 1991, he was an Assistant Professor with the Department of Electrical Engineering at Worcester Polytechnic Institute, Worcester, MA. From 1991 to 1996, he was an Associate Professor with the Department of Electrical and Computer Engineering at State University of New York, Buffalo. Since September 1996, he has been a Senior Research Scientist with the US Army Research Laboratory, Adelphi, MD, working on image processing and automatic target recognition. He has served as an Associate Editor for the IEEE Transactions on Image Processing, the IEEE Transactions on Circuits, Systems and Video Technology, and the IEEE Transactions on Neural Networks. His current research interests are in hyperspectral imaging, automatic target recognition, statistical machine learning theory, robotics, and neural networks applications to image processing. He is also a Fellow of ARL, SPIE, and IEEE.

**Yanning Zhang** received the B.S. degree from Dalian University of Science and Engineering, Dalian, China, in 1988, and the M.E. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 1993 and 1996, respectively. She is currently a Professor and an Associate Dean of School of Computer Science and Technology, Northwestern Polytechnical University. She is also the Organization Chair of the Ninth Asian Conference on Computer Vision (ACCV 2009). Her research interests include signal and image processing, computer vision, and pattern recognition.

**Thomas S. Huang** received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge. He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, and Research Professor at the Coordinated Science Laboratory, and at the Beckman Institute for Advanced Science he is Technology and Co-Chair of the Institute's major research theme "Human Computer Intelligent Interaction". His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 21 books, and over 600 papers in network theory, digital filtering, image processing, and computer vision. Dr. Huang is a Member of the National Academy of Engineering; a Member of the Academia Sinica, Republic of China; a Foreign Member of the Chinese Academies of Engineering and Sciences; and a Fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of America. Among his many honors and awards: Honda Lifetime Achievement Award, IEEE Jack Kilby Signal Processing Medal, and the KS Fu Prize of the Int. Asso. for Pattern Recognition.

Please cite this article as: H. Zhang, et al., Joint dynamic sparse representation for multi-view face recognition, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.09.009