REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
05-10-2012	Conference Proceeding	-

4. TITLE AND SUBTITLE
Multi-observation visual recognition via joint dynamic sparse representation

5a. CONTRACT NUMBER
W911NF-09-1-0383

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

6. AUTHORS
Haichao Zhang, Nasser M. Nasrabadi, Yanning Zhang, Thomas S. Huang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
William Marsh Rice University
Office of Sponsored Research
William Marsh Rice University
Houston, TX 77005 -

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
56177-CS-MUR.84

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
We address the problem of visual recognition from multiple observations of the same physical object, which can be generated under different conditions, such as frames at different time instances or snapshots from different viewpoints. We formulate the multi-observation visual recognition task as a joint sparse representation model and take advantage of the correlations among the multiple observations for classification using a novel joint dynamic sparsity prior. The proposed joint dynamic sparsity prior promotes shared joint sparsity pattern among the multiple

15. SUBJECT TERMS
Face , Face recognition , Heuristic algorithms , Joints , Training , Vectors , Visualization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Richard Baraniuk
UU	UU	UU	UU		19b. TELEPHONE NUMBER 713-348-5132

Report Title

Multi-observation visual recognition via joint dynamic sparse representation

ABSTRACT

We address the problem of visual recognition from multiple observations of the same physical object, which can be generated under different conditions, such as frames at different time instances or snapshots from different viewpoints. We formulate the multi-observation visual recognition task as a joint sparse representation model and take advantage of the correlations among the multiple observations for classification using a novel joint dynamic sparsity prior. The proposed joint dynamic sparsity prior promotes shared joint sparsity pattern among the multiple sparse representation vectors at class-level, while allowing distinct sparsity patterns at atom-level within each class in order to facilitate a flexible representation. The proposed method can handle both homogenous as well as heterogenous data within the same framework. Extensive experiments on various visual classification tasks including face recognition and generic object classification demonstrate that the proposed method outperforms existing state-of-the-art methods.

Conference Name: 2011 IEEE International Conference on Computer Vision (ICCV)

Conference Date: November 05, 2011

Multi-observation Visual Recognition via Joint Dynamic Sparse Representation

Haichao Zhang[†‡], Nasser M. Nasrabadi[§], Yanning Zhang[†] and Thomas S. Huang[‡]
[†] School of Computer Science, Northwestern Polytechnical University, Xi'an China
[‡] Beckman Institute, University of Illinois at Urbana-Champaign, IL USA
[§] U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD USA

Abstract

We address the problem of visual recognition from multiple observations of the same physical object, which can be generated under different conditions, such as frames at different time instances or snapshots from different viewpoints. We formulate the multi-observation visual recognition task as a joint sparse representation model and take advantage of the correlations among the multiple observations for classification using a novel joint dynamic sparsity prior. The proposed joint dynamic sparsity prior promotes shared joint sparsity pattern among the multiple sparse representation vectors at class-level, while allowing distinct sparsity patterns at atom-level within each class in order to facilitate a flexible representation. The proposed method can handle both homogenous as well as heterogenous data within the same framework. Extensive experiments on various visual classification tasks including face recognition and generic object classification demonstrate that the proposed method outperforms existing state-of-the-art methods.

1. Introduction

Recent dramatic increase in different kinds of visual data has created a surge in demand for effective processing and analysis algorithms. For instance, a video camera can generate multiple observations of the same object at different time instances; a camera network can capture the same subject from different viewpoints; systems with heterogenous sensors (*e.g.*, visible light cameras, inferred cameras and laser range finders) can generate heterogenous visual data for the same physical object. All these scenarios pose great challenges to the existing data processing techniques and require new schemes for effective data processing. In particular, object recognition and classification from multiple observations are interesting and are of great use for numerous applications (*e.g.*, surveillance, law enforcement). However, most existing techniques are designed for single observation based classification, which are clearly not optimal due to the failure of exploiting the correlations among the multiple observations of the same physical object.

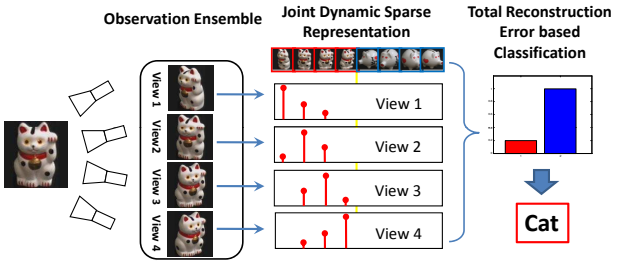


Figure 1. Illustration of the joint dynamic sparse representation based multi-observation recognition framework. The observation ensemble contains multiple observations generated under different conditions such as viewpoints. Each observation can be sparsely represented by potentially different training images from the same class, thus the sparse representation vectors share the same sparsity pattern at class level but distinct at atom level. Classification is achieved via the total reconstruction error of all the observations.

In this paper, we propose a novel Joint Dynamic Sparse Representation based Classification method (JDSRC) for multi-observation based visual recognition. The problem of recovering the sparse linear representation of a single query datum with respect to a set of reference datum (dictionary) has received wide interest recently in image processing, computer vision and pattern recognition communities [3, 10]. Recently, extensions on recovering the sparse representations of multiple query data samples jointly have been investigated and applied to multi-task visual recognition problem in [12], where the multiple tasks (features) are assumed to have the same sparsity pattern in their sparse representation vectors. The proposed JDSRC method exploits the correlations among the multiple observations using a novel joint dynamic sparsity prior to improve the performance of a recognition system, with the assumption that the sparse representation vectors of multiple observations have the same sparsity pattern at class level, but not necessarily at atom level, thus the proposed algorithm can not only exploit the correlations among the observations but is also more flexible than the same atom-level sparisty pattern assumption. Moreover, the JDSRC method is very gen-

2011 IEEE International Conference on Computer Vision
978-1-4577-1102-2/11/$26.00

eral, and can handle both homogenous and heterogenous data within the same framework. Taking multi-view object recognition as an example, Figure 1 depicts and motivates our JDSRC method. Given a set of test observations from different viewpoints for a given object "cat", we first perform 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 describing the same physical object "cat", the recovered sparse representation vectors tend to have the same sparsity pattern at class-level, ideally with non-zeros coefficients only associated with images of "cat" in the dictionary; on the other hand, since the multiple observations are captured from different viewpoints, the atom-level sparsity patterns of the representation vectors are not necessarily the same, tending to have non-zero coefficients associated with training images of similar viewpoints, as depicted in Figure 1. We term the property that multiple sparse representation vectors with shared sparsity pattern at class-level but not necessarily at atom-level as *joint dynamic sparsity*. Using this property, the proposed JDSRC method can achieve several significant goals: (1) it combines the information from each observation for discrimination during the joint sparse recovery process rather than in post-processing, thus can potentially avoid the risk of making erroneous decision for each observation when treated independently; (2) it exploits the correlations among all the observations and can handle both homogenous and heterogenous tasks; (3) the joint dynamic sparsity model adopted in JDSRC enables more flexible and adaptive atom selection for joint sparse representation, thus is more powerful. The rest of this paper is organized as follows. In Section 2, we review some related works briefly. We introduce the JDSRC model and present an efficient algorithm for solving it in Section 3. Experiments are carried out on various datasets in Section 4. We conclude this paper in Section 5.

2. Related Works

We will first review the sparse representation based method for single observation based classification and then discuss its recent extension to multiple observations.

2.1. Single Observation based Classification via Sparse Representation

Recently, a Sparse Representation based Classification (SRC) method for single image based face recognition is developed in [11]. This method casts the task of face recognition as one of classifying between linear regression models via sparse representation. It is based on the simple assumption that a new test sample $\mathbf{y}$ from the c-th class lies in the same subspace as the training samples (atoms) of the same class $\mathbf{A}_c = [\mathbf{a}_{c,1}, \mathbf{a}_{c,2}, \cdots]$, thus can be well represented by a linear combination of the training samples from $\mathbf{A}_c$:

$$\mathbf{y} = x_{c,1}\mathbf{a}_{c,1} + x_{c,2}\mathbf{a}_{c,2} + \cdots = \mathbf{A}_c\mathbf{x}_c. \tag{1}$$

As the class label of the test image $\mathbf{y}$ is unknown, we can recover the representation vector $\mathbf{x}$ for $\mathbf{y}$ with respect to the whole training set $\mathbf{A}$, which should be sparse by assumption, thus naturally leading to a sparse representation problem over $\mathbf{A}$ [11]:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon, \end{aligned} \tag{2}$$

where ϵ is the reconstruction error parameter, $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_C] \in \mathbb{R}^{d\times N}$ is the dictionary collecting training samples from all C classes and $\mathbf{x} = [\mathbf{x}_1^\top, \mathbf{x}_2^\top, \cdots, \mathbf{x}_C^\top]^\top \in \mathbb{R}^N$ is the representation vector in terms of $\mathbf{A}$. $N = \sum_{c=1}^{C} N_c$ is the total number of training samples. After recovering $\hat{\mathbf{x}}$, the class label for $\mathbf{y}$ is determined based on the minimum reconstruction error criteria by projecting the test sample onto each class as:

$$\hat{c} = \arg\min_c \|\mathbf{y} - \mathbf{A}\delta_c(\hat{\mathbf{x}})\|_2^2, \tag{3}$$

where $\delta_c(\cdot)$ is a vector operator keeping the elements corresponding to the c-th class while setting all others as zero.

2.2. Classification via Joint Sparse Representation for Multiple Observations

In presence of multiple observations, applying SRC for each observation separately is clearly sub-optimal due to the failure of exploiting the correlations among the multiple observations. It should be more robust to perform sparse representation simultaneously for all the observations, while combining the information from all of them during sparse recovery by applying joint constraints to their sparse representation vectors. Recently, several extensions have been made to generalize SRC to handle multiple observations. Denoting the dictionary associated with the k-th observation $\mathbf{y}^k$ as $\mathbf{A}^k$ (also referred to as observation-dictionary), $k \in \{1, 2, \cdots, K\}$, [12] proposed a Multi-Task Joint Sparse Representation Classification (MTJSRC) method for multiple feature based classification:

$$\begin{aligned} \hat{\mathbf{X}} &= \arg\min_{\mathbf{X}} \frac{1}{2}\sum_{k=1}^{K} \|\mathbf{y}^k - \sum_{c=1}^{C} \mathbf{A}_c^k\mathbf{x}_c^k\|_2^2 + \lambda\sum_{c=1}^{C}\|\mathbf{x}_c\|_2 \\ &= \arg\min_{\mathbf{X}} \frac{1}{2}\sum_{k=1}^{K} \|\mathbf{y}^k - \mathbf{A}^k\mathbf{x}^k\|_2^2 + \lambda\sum_{c=1}^{C}\|\mathbf{x}_c\|_2, \end{aligned} \tag{4}$$

where $\mathbf{x}_c = [\mathbf{x}_c^{1\top}, \cdots, \mathbf{x}_c^{K\top}]^\top \in \mathbb{R}^{KN_c}$ is the collection of the representation vectors associated with class c across all the K observations/features. $\mathbf{A}^k = [\mathbf{A}_1^k, \mathbf{A}_2^k, \cdots, \mathbf{A}_C^k] \in \mathbb{R}^{d_k\times N}$ denotes the dictionary for the k-th observation and

$\mathbf{x}^k = [\mathbf{x}_1^{k\top}, \mathbf{x}_2^{k\top}, \cdots, \mathbf{x}_C^{k\top}]^\top \in \mathbb{R}^N$ is the associated representation vector. $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^K] \in \mathbb{R}^{N\times K}$ is the collection of the multiple sparse representation vectors. Using this model, the recovered sparse representation vectors will have the same sparsity pattern, not only at class-level, but at atom-level as well. The classification decision is made as the class which gives the lowest reconstruction error accumulated over all the K observations:

$$\hat{c} = \arg\min_c \sum_{k=1}^{K} w^k \|\mathbf{y}^k - \mathbf{A}^k \delta_c(\mathbf{x}^k)\|_2^2, \tag{5}$$

where w^k is the weight reflecting the confidence in the k-th observation which can be learned from data.

3. Joint Dynamic Sparse Representation based Multi-observation Recognition

3.1. Problem Formulation

The MTJSRC method [12] generalizes SRC to multiple observations by assuming all the observations will share the same set of selected atoms for sparse representation (Figure 2 (b)), which is reasonable in the case of multiple features from the same datum. However, in more general multiple observations cases, due to the variation of observation conditions, *e.g.*, viewpoints, each observation may be better represented by a different set of atoms from the same class, as illustrated in Figure 1, thus assuming the observations from different viewpoints can be represented by the same set of training samples is inappropriate. Rather, 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 shown in Figure 2 (c). One of the key ingredients in our JDSRC model for promoting joint dynamic sparsity is the *dynamic active set*. A dynamic active set $\mathbf{g}_s \in \mathbb{R}^K$ refers to the indices of a set of coefficients corresponding to the same class in the coefficient matrix $\mathbf{X}$, which are activated jointly during sparse representation of multiple observations. Each dynamic active set $\mathbf{g}_s$ contains one and only one index for each column of $\mathbf{X}$, where $\mathbf{g}_s(k)$ is for the k-th column of $\mathbf{X}$, as shown in Figure 2 (c).

We formulate our JDSRC 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 observations 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 observations), we apply ℓ_2-norm

(a) (b) (c)

Figure 2. Illustration of different sparsity models for coefficient matrix $\mathbf{X}$. Each column denotes a representation vector and each squared block denotes a representation coefficient value in the corresponding representation vector. A white block denotes zero entry value. Colored blocks denote non-zeros values. (a) separate sparse representation: the sparse representation vectors may be quite different due to the separate recovery process. (b) joint sparsity: sparse coefficient vectors share the same patterns (selecting the same atoms), but with different coefficient values. (c) joint dynamic sparsity: the sparse coefficient vectors select different atoms within each class-dictionary to represent each of the observations.

over each dynamic active set; to promote sparsity, *i.e.*, to allow a small number of dynamic active sets to be involved in joint sparse representation, we apply ℓ_0-norm across the ℓ_2-norm of the dynamic active sets. Therefore, we arrive at the following joint dynamic sparsity promoting term:

$$\|\mathbf{X}\|_G = \left\| \left[\|\mathbf{x}_{\mathbf{g}_1}\|_2, \|\mathbf{x}_{\mathbf{g}_2}\|_2, \cdots \right] \right\|_0, \tag{6}$$

where $\mathbf{x}_{\mathbf{g}_s}$ denotes the vector formed as the collection of the coefficients associated with the s-th dynamic active set $\mathbf{g}_s$: $\mathbf{x}_{\mathbf{g}_s} = \mathbf{X}(\mathbf{g}_s) = [\mathbf{X}(\mathbf{g}_s(1), 1), \mathbf{X}(\mathbf{g}_s(2), 2), \cdots, \mathbf{X}(\mathbf{g}_s(K), K)]^\top \in \mathbb{R}^K$. To recover the sparse representation coefficient matrix $\mathbf{X}$ with joint dynamic sparse property for the multiple observations $\{\mathbf{y}^k\}_{k=1}^K$, we propose the following Joint Dynamic Sparse Representation (JDSR) model:

$$\begin{aligned} \hat{\mathbf{X}} = \arg\min_{\mathbf{X}} & \sum_{k=1}^{K} \|\mathbf{y}^k - \mathbf{A}^k \mathbf{x}^k\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{X}\|_G \le S, \end{aligned} \tag{7}$$

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

- ℓ_2-norm is applied over each dynamic active set, thus allowing to combine the cues from all the observations during joint sparse representation; moreover, allowing each dynamic active set to be adaptive within the same class is both more flexible and reasonable due to the fact that the multiple observations are different measurements of the same physical object;

- ℓ_0-norm is applied across the dynamic active sets, thus encouraging the selection of the most parsimonious

and representative dynamic active sets, which promotes joint sparsity pattern shared at class-level while allows the within-class sparsity patterns to be distinct to facilitate the selection of the most representative atoms for each observation class-wise.

3.2. An Efficient Algorithm for Joint Dynamic Sparse Representation

The JDSR model (7) is very challenging to solve due to the co-existence of ℓ_0-norm and joint dynamic sparse property. We propose to solve (7) with a greedy JDSR algorithm as detailed in Algorithm 1. The proposed JDSR algorithm has a similar algorithmic structure as SOMP [9] and CoSOMP [2], which includes the following general steps: (i) select new candidates based on the current residue; (ii) merge the newly selected candidate set with previous 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 [2]. 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 for the rows). The major difference of JDSR with CoSOMP [2] lies in the atom selection criteria used in steps (i) and (iv) of Algorithm 1, which is detailed in the sequel.

At each iteration of JDSR (step (i) and (iv)), given a coefficient matrix $\mathbf{Z} \in \mathbb{R}^{N\times K}$, 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:

$$\begin{aligned} \hat{\mathbf{Z}}_L &= \arg\min_{\mathbf{Z}\in\mathbb{R}^{N\times K}} \|\mathbf{Z}-\mathbf{Z}_L\|_{\mathcal{F}} \\ \text{s.t} \quad & \|\mathbf{Z}_L\|_G \le L. \end{aligned} \tag{8}$$

The solution to (8) can be obtained by a procedure called the Joint Dynamic Sparsity mapping (JDS mapping):

$$\mathbf{I}_L = \mathbb{P}_{\text{JDS}}(\mathbf{Z}, L), \tag{9}$$

which gives the index matrix $\mathbf{I}_L \in \mathbb{R}^{L\times K}$ containing the top-L dynamic active sets for all the K observations, as detailed in Algorithm 2. In 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 observation; (ii) combine the maximum absolute coefficients across the observations for each class as the total response; (iii) select the dynamic active set as the one which 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 in each iteration of Algorithm 1 and this iteration process is repeated on the residue until certain conditions are satisfied [2, 9].

Algorithm 1: Joint Dynamic Sparse Representation (JDSR) based Classification (JDSRC).

Input: observation set $\{\mathbf{y}^k\}_{k=1}^K$, dictionary set $\{\mathbf{A}^k\}_{k=1}^K$, sparsity level S, observation number K

Output: class label $\hat{c}$

while *stopping criteria false* **do**

- $\mathbf{E}(:,k) = \mathbf{A}^{k\top}\mathbf{r}^k, \forall k = 1,2,\cdots,K;$
- % **(i)** atom selection via joint dynamic sparse mapping
- $\mathbf{I}_{\text{new}} \leftarrow \mathbb{P}_{\text{JDS}}(\mathbf{E}, 2S)$;
- $\mathbf{I} \leftarrow [\mathbf{I}^\top, \mathbf{I}_{\text{new}}^\top]^\top$ % **(ii)** index matrix updating;
- % **(iii)** representation coefficients updating
- **for** $k = 1,2,\cdots,K$ **do**
  - $\mathbf{i} \leftarrow \mathbf{I}(:,k);$
  - $\mathbf{C}(\mathbf{i},k) \leftarrow (\mathbf{A}^k(:,\mathbf{i})^\top\mathbf{A}^k(:,\mathbf{i}))^{-1}\mathbf{A}^k(:,\mathbf{i})^\top\mathbf{y}^k;$
- % **(iv)** atom pruning via joint dynamic sparse mapping
- $\mathbf{I} \leftarrow \mathbb{P}_{\text{JDS}}(\mathbf{C}, S)$;
- $\mathbf{X} \leftarrow \mathbf{0};$
- **for** $k = 1,2,\cdots,K$ **do**
  - $\mathbf{i} \leftarrow \mathbf{I}(:,k), \quad \mathbf{X}(\mathbf{i},k) \leftarrow \mathbf{C}(\mathbf{i},k);$
  - $\mathbf{r}^k = \mathbf{A}^k\mathbf{X}(:,k) - \mathbf{y}^k$ % **(v)** residue updating;

for $k = 1,2,\cdots,K$ **do**

- $\mathbf{i} \leftarrow \mathbf{I}(:,k);$
- $\mathbf{X}(\mathbf{i},k) \leftarrow (\mathbf{A}^k(:,\mathbf{i})^\top\mathbf{A}^k(:,\mathbf{i}))^{-1}\mathbf{A}^k(:,\mathbf{i})^\top\mathbf{y}^k;$

$\hat{\mathbf{y}}_c^k = \mathbf{A}^k\delta_c(\mathbf{X}(:,k))$ % reconstruction;

$e_c = \sum_{k=1}^K w^k\|\mathbf{y}_c^k - \hat{\mathbf{y}}_c^k\|_2^2$ % total reconstruction error;

$\hat{c} = \arg\min_c e_c$ % class label estimation.

3.3. Classification Rule

After recovering the sparse representations matrix $\hat{\mathbf{X}} = [\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \cdots, \hat{\mathbf{x}}^K]$ for all the observations $\{\mathbf{y}^k\}_{k=1}^K$ of the same physical object via JDSR, we make a decision on the class label jointly for all the observations based on $\hat{\mathbf{X}}$, which is achieved via the total reconstruction error criteria as:

$$\hat{c} = \arg\min_c \sum_{k=1}^K w^k\|\mathbf{y}^k - \mathbf{A}^k\delta_c(\hat{\mathbf{x}}^k)\|_2^2, \tag{10}$$

where $\{w^k\}_{k=1}^K$ are the confidence weights for the observations. Using total reconstruction error for classification, we again combine the cues from all the observations. The overall procedures of JDSRC are summarized in Algorithm 1.

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

Input: coefficient matrix $\mathbf{Z}$, desired number of dynamic active sets L, label vector $\mathbf{L}$ for atoms in the dictionary, number of classes C, number of observations K

Output: index matrix $\mathbf{I}_L$ for the top-L dynamic active sets

Initialize: $\mathbf{I}_L \leftarrow \emptyset$ % initialize the index matrix as empty;

for $l = 1, 2, \cdots, L$ **do**

  for $c = 1, 2, \cdots, C$ **do**

    $\mathbf{c} \leftarrow \text{find}(\mathbf{L}, c)$ % get the index vector for the c-th class;

    for $k = 1, 2, \cdots, K$ **do**

      % **(i)** find the maximum absolute value v and its index t for the c-th class, k-th observation

      $[v, t] \leftarrow \max(|\mathbf{Z}(\mathbf{c}, k)|)$;

      $\mathbf{V}(c, k) \leftarrow v, \quad \tilde{\mathbf{I}}(c, k) \leftarrow \mathbf{c}(t)$;

    % **(ii)** combine the max-coefficients for each class

    $\mathbf{s}(c) \leftarrow \sqrt{\sum_{k=1}^{K} \mathbf{V}(c, k)^2}$;

  $[\hat{v}, \hat{t}] = \max(\mathbf{s})$ % **(iii)** find the best cluster of atoms belonging to the same class across all the classes;

  $\mathbf{I}_L(l, :) = \tilde{\mathbf{I}}(\hat{t}, :), \quad \mathbf{Z}(\tilde{\mathbf{I}}(\hat{t}, :)) \leftarrow \mathbf{0}^\top$.

4. Experiment Results

In this section, we evaluate the performance of the proposed JDSRC method on several visual classification applications. Specifically, we carry out experiments on multi-region based face recognition, multi-instance based face recognition and multi-view visual recognition. To verify the effectiveness of the proposed method, we compare the proposed method with several *state-of-the-art* methods, including: SRC [11], MTJSRC [12], Mutual Subspace Method (MSM) [4] and Affine Hull (AFH) method for set based classification [1]. The weight for each observation can be learned via a learning procedure. For the applications demonstrated in the sequel, all the observations can be regarded as equally important for classification, thus all the weights are set to be equal without loss of generality.

4.1. Multi-Region Face Recognition

Local region/patch based face recognition methods have been proven to be effective in literature. In this subsection, we treat each region from a face image as a single observation. Eight regions ($K = 8$) are manually selected in this experiment as illustrated in Figure 3 (a): (left, right) brows, eyes, cheeks, nose and mouth, thus inducing a heterogenous recognition task. Since different observations have different properties and they can not be matched with each other, dedicated observation-dictionaries are required for each region. The k-th observation-dictionary $\mathbf{A}^k$ is constructed from the corresponding k-th region of all the training images.

4.1.1 Holistic SRC, Separate-region SRC and Multi-region Joint Dynamic Sparse Representation

In this illustrative experiment, we compare the recovered sparse representation vector(s) using: SRC on the holistic face, SRC on each region separately and the proposed JDSRC method. For illustration, we select 5 classes from the Extended Yale B dataset [5] where each class contains 32 gallery faces. Representative faces for each class are shown in Figure 3 (d). The probe face is shown in Figure 3 (b), which belongs to class 3. The probe face is under extremely low-illumination condition, thus for better visualization, an enhanced version of the probe face is shown in Figure 3 (c). We infer the label for the probe face with the holistic sparse representation, separate sparse representation on each region as well as the proposed JDSRC method. The results are shown in Figure 4 $\sim$ 6. For holistic-SRC, the recovered sparse representation vector as well as the reconstruction errors are shown in Figure 4. As can be seen, this method tends to predict the probe face as from class 1, which is incorrect. For separate-SRC, as each region is treated independently, the sparse representation vectors for different regions are quite distinct, as shown in Figure 5, thus although some regions prefer the correct label, overall it makes an incorrect decision which is again class 1. The proposed JDSRC method can combine the cues from all the 8 regions during sparse representation by matching each region with the corresponding region of different gallery images of the same person, thus providing a more robust class label prediction. As shown in Figure 6, the recovered sparse coefficients are mostly concentrated at the correct class (class 3, black) while the within-class non-zero supports are different, indicating each region matches with different gallery images of the same person, thus is more flexible. The final reconstruction error achieves a minimum at class 3, which is the correct label for the probe face image.

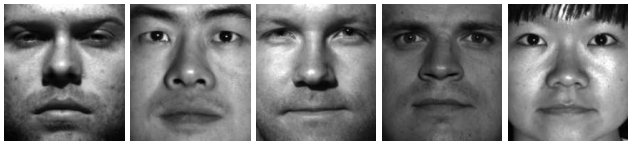
4.1.2 Multi-Region Face Recognition

In this subsection, we compare the recognition performance of JDSRC method with SRC [11] and MTJSRC [12] on the Extended Yale B dataset [5] (192×168 pixels). The partitions depicted in Figure 3 (a) is used. We follow the experimental setups in [11] for a fair comparison. Specifically, all the 2414 frontal views of 38 individuals are used and are resized to 24×21. Half of the images randomly sampled from the whole database are used for training and the rest for testing. We set sparisty level as $S = 25$. Recognition rates for different algorithms under this setting are summarized in Table 1. As can be seen from Table 1, the proposed JDSRC method clearly outperforms Nearest Neigh-

(a) Partition

(b) Probe

(c) Enhanced



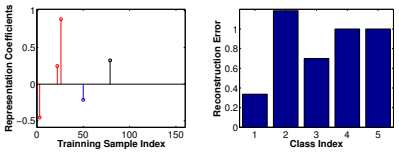
(d) Representative gallery faces

Figure 3. Face images from Extended Yale B dataset. (a) the 8 selected regions used in our experiments: (left, right) brows, eyes, cheeks, nose and mouth. Some face images used in Subsection 4.1.1: (b) original probe face, (c) enhanced probe face for visualization, (d) representative faces from the gallery set.

bor (NN), linear SVM (SVM) as well as holistic SRC [11] and MTJSRC [12] on multiple regions. The performances of different algorithms under different number of training samples are depicted in Figure 7 (a), which demonstrates that the proposed JDSRC method outperforms holistic SRC constantly. We also examine the performance of each algorithm under different image sizes with down-sampling factor of $r \in \{24, 16, 8, 6, 4\}$. The performances of different algorithms are shown in Figure 7 (b). As can be seen from Figure 7 (b), by decreasing the down-sampling factor (*i.e.* increasing the dimensionality of features), the recognition rates increase for all the algorithms. The behaviors of different algorithms are, however, different. The best accuracy for NN under the highest feature dimension is still lower than that of all the other algorithms under the lowest feature dimension. SVM achieves a relatively low accuracy at the lowest feature dimension, and improves the performance quickly as the dimension increases. SRC method can achieve a relatively higher accuracy at the lowest feature dimension, but its performance improves slowly as the dimensionality increases. The proposed JDSRC method, on the other hand, also achieves a high recognition accuracy at the lowest feature dimension, which is approximately the same as SRC. Moreover, as the feature dimension increases, JDSRC increases its performance quickly and achieves a recognition accuracy of over 99% when the feature dimension is larger than 504, which clearly outperforms SRC.

4.2. Multi-Instance Face Recognition

In this experiment, we consider the scenario of having multiple instances of a subject for classification, as in the case of multiple frames generated from video cameras which is a typical scenario in surveillance. In such an unconstrained environment, the captured face images may have large intra-class pose variations. UMIST face database



(a) (b)

Figure 4. Holistic face sparse representation. (a) sparse representation coefficient plot (b) reconstruction error bar plot

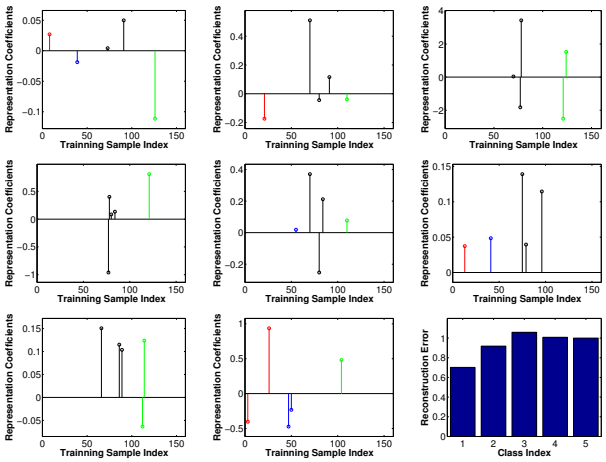


Figure 5. Separate regions based face sparse representation: 8 sparse coefficients plots and reconstruction error bar plot.

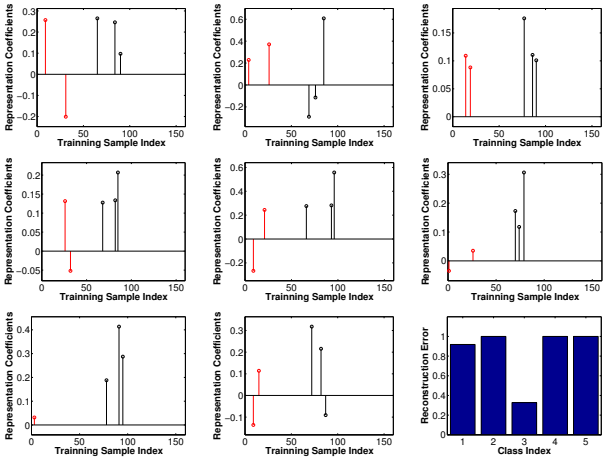
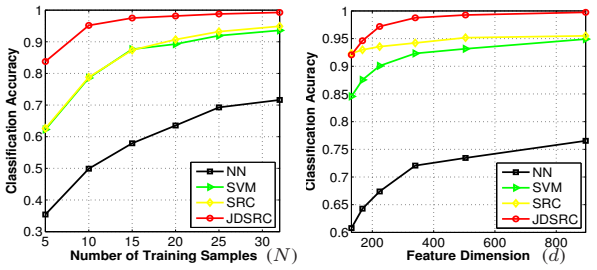


Figure 6. Joint dynamic sparse representation: 8 sparse coefficients plots and reconstruction error bar plot.

Table 1. Multi-region face recognition accuracy (%) on the Extend Yale B with feature dimension $d = 504$.

Algorithm	Recognition Accuracy
NN	59.85
SVM	93.59
SRC [11]	97.10
MTJSRC [12]	98.05
JDSRC	**99.34**



(a) (b)

Figure 7. Recognition accuracy plots on Extended Yale B. (a) recognition accuracy under different number of training samples; (b) recognition accuracy under different feature dimensions.

Figure 8. Sample images from UMIST database for a single subject with varying poses.

Table 2. Multi-instance face recognition accuracy (%) on UMIST.

Algorithm	2 Views	3 Views	4 Views	Avg.
MSM [4]	93.5	95.0	96.5	95.0
AFH [1]	93.0	95.5	97.0	95.2
MTJSRC [12]	94.5	95.5	**98.0**	96.0
JDSRC	**95.5**	**97.5**	**98.0**	**97.0**

is used in this experiment, consisting of 564 images of 20 individuals (mixed race/gender) [7]. Each individual is shown in a range of poses from profile to frontal views, as shown in Figure 8. We randomly select 10 images for each individual to construct the observation-dictionary, which is shared by all the observations. For testing, we regard each image as a single observation and carry out experiments under different number of observations ($K = \{2, 3, 4\}$) selected randomly from the rest of the database for each individual. We set the sparsity level as $S = 5$. Experiment results are summarized in Table 2. As the multiple observations are not likely to have exactly the same pose, they are more likely to match with different set of training faces of the same subject in the gallery, which can not be handled well by the MTJSRC method, as also revealed by the results in Table 2. As can be seen, the proposed JDSRC method performs better than the other methods in this experiment.

4.3. Multi-View Visual Recognition

We apply JDSRC to visual recognition from multiple view images. First, we use ALOI dataset for experiment,

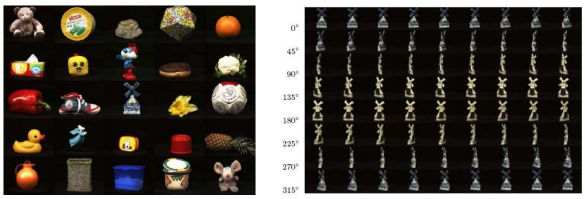


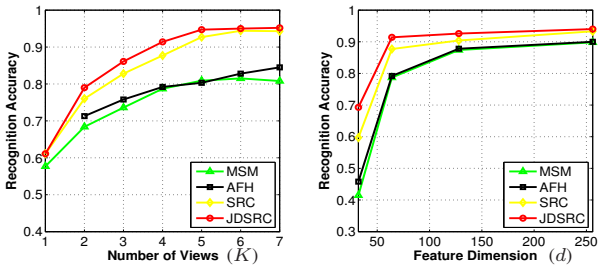
Figure 9. ALOI database. Left: sample images from ALOI database. Right: 72 different viewpoints for a specific object.

Table 3. Multi-view object classification accuracy (%) on ALOI.

Algorithm	2 Views	4 Views	6 Views	Avg.
MSM [4]	**97.1**	97.1	**100.0**	98.1
AFH [1]	94.3	71.4	57.1	74.3
MTJSRC [12]	90.0	94.3	98.6	94.3
JDSRC	**97.1**	**98.6**	**100.0**	**98.6**

which is a image collection of 1000 small objects [6], with systematically varied viewing angle, illumination angle, and illumination color for each object. Sample images and illustration of the 72 viewpoints for each object are depicted in Figure 9. In this experiment, we select a subset of 70 classes for computational consideration in algorithm evaluation. Images from 6 viewpoints corresponding to view angles $\Theta_{\text{train}} = \{0°, 60°, 120°, 180°, 240°, 300°\}$ are used for training. We test the performance of different algorithms by randomly selecting different number of views $K = \{2, 4, 6\}$ from the remaining viewpoints for each object. Therefore, training and testing images are recorded from different viewpoints. We set sparsity level as $S = 5$. The results are summarized in Table 3, which further verify the effectiveness of the proposed JDSRC method compared with the other methods and demonstrate the applicability of the proposed method on general visual classification tasks.

We further apply our JDSRC method to multi-view face recognition using CMU Multi-PIE database [8], which contains a large number of face images under different illuminations, viewpoints and expressions, up to 4 sessions over the span of several months. Subjects were imaged under 13 cameras at head height, spaced at 15° intervals and 20 illumination conditions. In our experiment, the face regions for all poses are extracted manually and are resized to 45×35. We choose the first 50 classes which are present in all the 4 sessions for experiment. Due to the symmetric property of human faces, we consider only 7 different poses with view angles $\Theta = \{0°, 15°, 30°, 45°, 60°, 75°, 90°\}$. 4 different view angles $\Theta_{\text{train}} = \{0°, 30°, 60°, 90°\}$ from Session 1 are used for training while all the 7 different view angles in Θ from the Session $2 \sim 4$ are used for testing. This is a more realistic setting in the sense that the data sets used for training and testing are collected separately and even not all the viewpoints in the testing sets are available for training. Images with expressions are not used in our experimental



(a) (b)

Figure 10. Recognition rate under different (a) number of views ($d = 128$) and (b) feature dimensions ($K = 4$).

evaluation. We set sparsity level as $S = 5$ and use random projection for dimensionality reduction [11].

To generate a test sample with K views, we first randomly select a subject $c \in \{1, 2, \cdots, 50\}$ from the test set and then randomly select $K \in \{1, 2, \cdots, 7\}$ different views imaged at the same time instance for subject c. 1000 test samples are generated with this scheme for testing. For SRC, sparse representation procedure is performed for each view separately and then a single decision is made based on the recovered coefficient vectors using (10). The MTJSRC method [12] is not compared in this experiment, as the same sparsity pattern assumption it makes is improper for this task, thus limiting its performance (as can also be observed in Table 3). The recognition results on Session 2 are shown in Figure 10 (a). It is demonstrated that the multi-view based methods ($K > 1$) outperform their single-view counterparts ($K = 1$) by a large margin, indicating the advantage of using multiple views in face recognition. Furthermore, it is noted that the performance of all the algorithms improves as the number of views is increased and the proposed method outperforms all the other methods under all different number of views. We also examine the effects of data (feature) dimensionality d on recognition rate. The test samples are generated using Θ with $K = 4$. We vary the data dimensionality in the range of $d \in \{32, 64, 128, 256\}$ and show in Figure 10 (b) the plots of the performances for all the algorithms on Session 2 data set. It is shown that the proposed JDSRC method performs the best under all the examined dimensionality of features. Finally, we evaluate the performance of all the algorithms on different sessions from Multi-PIE. The recognition results on Session $2 \sim 4$ data set with $d = 128$ are summarized in Table 4. It is demonstrated that the proposed JDSRC method outperforms all the other algorithms on different test sessions.

5. Conclusion

A novel joint dynamic sparse representation based visual recognition method is presented in this paper. This method inherits the robustness of the sparse representation based classification method while also has the advantage of exploiting the correlations among the multiple observations. Moreover, the novel joint dynamic sparsity model allows more flexible atom selection for joint sparse representation, which facilitates recognition. Experimental results of the proposed method compared with *state-of-the-art* methods on various visual recognition tasks verified the effectiveness of the proposed method. For future work, we would like to address theoretical aspects of the proposed method. Also, we would like to further explore other applications of the proposed method, such as multi-modal visual classification.

Table 4. Multi-view face recognition rate (%) on different test sessions of CMU Multi-PIE database ($C = 50, d = 128, K = 4$).

Algorithm	Session 2	Session 3	Session 4
MSM [4]	87.4	81.0	76.9
AFH [1]	87.8	82.5	78.3
SRC [11]	90.4	88.5	85.6
JDSRC	**92.6**	**91.6**	**86.7**

Acknowledgements This work is supported by NSF (60872145, 60903126), National High-Tech.(2009AA01Z315), Postdoctoral Science Foundation (20090451397, 201003685) and Cultivation Fund from Ministry of Education (708085) of China. This work is also supported by U.S. Army Research Laboratory and U.S. Army Research Office under grant number W911NF-09-1-0383.

References

[1] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *CVPR*, 2010. 5, 7, 8

[2] M. F. Duarte, V. Cevher, and R. G. Baraniuk. Model-based compressive sensing for signal ensembles. In *Proc. the 47rd Allerton Conf. on Communication, Control, and Computing*, 2009. 4

[3] M. Elad, M. Figueiredo, and Y. Ma. On the role of sparse and redundant representations in image processing. *Proc. IEEE*, 98(6):972–982, 2010. 1

[4] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. *Springer Tracts in Advanced Robotics*, 15:192–201, 2005. 5, 7, 8

[5] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE TPAMI*, 23(6):643–660, 2001. 5

[6] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The amsterdam library of object images. *Int. J. Comput. Vision*, 61(1):103–112, 2005. 7

[7] D. B. Graham and N. M. Allinson. Face recognition: From theory to applications. *NATO ASI Series F, Computer and Systems Sciences*. 7

[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vis. Comput.*, 28(5):807–813, 2010. 7

[9] J. A. Tropp, A. C. Gilbert, Martin, and J. Strauss. Algorithms for simultaneous sparse approximation. *EURASIP J. App. Signal Processing*, 2006. 4

[10] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proc. IEEE*, 98(6):1031–1044, 2010. 1

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 2009. 2, 5, 6, 8

[12] X.-T. Yuan and S. Yan. Visual classification with multi-task joint sparse representation. In *CVPR*, 2010. 1, 2, 3, 5, 6, 7, 8