Discriminative Representative Selection via Structure Sparsity

Baoxing Wang*, Qiyue Yin†, Shu Wu*†, Liang Wang† and Guiquan Liu‡
*School of Software and Engineering, University of Science and Technology of China
Email: newman97@mail.ustc.edu.cn
†Center for Research on Intelligent Perception and Computing (CRIPAC), NLPR, CASIA
Email: {qyyin, shu.wu, wangliang}@nlpr.ia.ac.cn
‡School of Computer Science and Technology, University of Science and Technology of China
Email: gqliu@ustc.edu.cn

***Abstract*—This paper focuses on the problem of finding a few representatives for a given dataset, which have both representation and discrimination ability. To solve this problem, we propose a novel algorithm, called Structure Sparsity based Discriminative Representative Selection (SSDRS), to find a representative subset of data points. The selected representative subset keeps the representation ability based on sparse representation models assuming that each data point can be expressed as a linear combination of those representatives. Meanwhile, we employ the Fisher discrimination criterion to make the coefficient matrix possess small within-class scatter but big between-class scatter, which leads to the discriminant ability of representatives. Since such a selected subset is representative and discriminative, it can be used to properly describe the entire dataset and achieve a good classification performance simultaneously. Experimental results in terms of video summarization and image classification indicate that our proposed algorithm outperforms the state-of-the-art methods.**

I. Introduction

In the community of pattern recognition and computer vision, a common problem is how to efficiently deal with large scale datasets, such as databases of images, videos and text documents. Since dealing with massive datasets is time-consuming and memory-consuming, many relevant methods have been developed such as dimensionality reduction and online learning. These methods can, to some degrees, deal with large scale datasets and improve the classification performance as well. But they generally need to take all data points into consideration.

To handle large scale datasets, representative selection [1] has recently been proposed, which is different from dimensionality reduction. Representative subset selection aims at finding a few representatives in the object space that can appropriately represent the whole dataset, thus the tasks such as clustering and classification can be done by using representatives instead of the entire dataset. Recent works [2, 3] indicate that classification algorithms can achieve a good performance if the representatives are informative enough about the given data. Many typical machine learning algorithms which are not originally designed for massive datasets can be applied to large scale datasets by using representatives.

To find a representative subset, several methods have been proposed [3, 4, 5] in recent years. Kmedoids [3], similar to Kmeans, is an unsupervised representative selection method which finds a number of data centers surrounded by other data points. These centroids are treated as representatives. Different from Kmedoids, the Affinity Propagation [5] algorithm finds some data centers, which are chosen as a condensed view of the whole dataset, by using a message passing algorithm when similarities between pairs of samples are given. By assuming data to be low-rank, some matrix factorization methods [1, 6, 7] can be used for representative selection. The Rank Revealing QR (RRQR) [1] algorithm aims at finding a proper permutation matrix that can generate the best conditioned submatrix and then obtain a few representatives. When a low-rank data matrix has missing entries, another matrix factorization method in [6] can be used to select a subset of samples by employing a greedy algorithm. Moreover, a nonnegative matrix factorization method [7] has been proposed to deal with the data matrix, which contains nonnegative entries. This method selects a few columns of the data matrix by using an l_1/l_∞ optimization.

Recently, the sparse model based representative subset selection methods [2] have been demonstrated to be effective in finding representatives. The Sparse Modeling Representative Selection (SMRS) method [2] finds a few representatives by solving a sparse multiple measurement vector problem, and those selected representatives are correlated to the nonzero rows of the coding coefficient matrix. Previous works are basically proposed under the assumption that data points are either lie in a low-dimensional space or have a natural cluster structure. Meanwhile, most of them only aim at finding a subset of data points that can properly describe the entire dataset. However, these methods seldom take the discriminant ability of representatives into consideration, which is very important to the classification task.

In this paper, we propose a novel representative subset selection algorithm, named **S**tructure **S**parsity based **D**iscriminative **R**epresentative **S**election (**SSDRS**), to select a subset of data points that have the representational capacity and discriminant ability at the same time. Our goal is achieved by minimizing the objective function which consists of reconstruction fidelity term and discrimination constraint term. Specially, the reconstruction fidelity term is based on the sparse representation

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.250

model, and the discrimination constraint term is achieved by imposing the Fisher discrimination criterion on the coefficient matrix so that those selected representatives belonging to the same class will have the smallest within class scatter and the representatives belonging to different classes will have the biggest between class scatter. Thus, representatives selected by our algorithm not only can well express the entire dataset but also can make representatives belonging to different classes are more differentiable. To the best of our knowledge, this is the first work that focuses on finding a representative subset which can simultaneously well describe the entire dataset and make representatives belonging to different classes discriminative. Experimental results with respect to two applications, i.e., video summarization and image classification, show that our proposed algorithm outperforms the state-of-the-art algorithms.

The rest of this paper is organized as follows. In Section 2, we briefly review the most related SMRS algorithm. Section 3 presents the formulation of the proposed SSDRS algorithm. In Section 4, experimental results and discussion are given. Finally, we conclude our work in Section 5.

II. SMRS Revisited

Our proposed algorithm is motivated by the Sparse Modeling Representative Selection method (SMRS) [2]. In order to put our work into context, we first briefly introduce the SMRS algorithm. SMRS is developed under the assumption that each data point can be expressed as a linear combination of a few representatives. The formulation of SMRS is provided as follows.

Given an $m \times N$ data matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N]$ which consists of N data points $\{\boldsymbol{y}_i \in \mathbb{R}^m\}_{i=1}^N$, SMRS tries to learn a compact coefficient matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{N \times N}$ by constraining the coefficient matrix $\boldsymbol{X}$ to be sparse. In this way, each data point can be well represented by minimizing the following objective function,

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}\|_F^2 \quad s.t. \quad \|\boldsymbol{X}\|_{q,0} \leq k, \mathbf{1}^T\boldsymbol{X} = \mathbf{1}^T, \quad (1)$$

where the mixed l_0/l_q norm is defined as $\|\boldsymbol{X}\|_{q,0} \triangleq \sum_{i=1}^N I(\|\boldsymbol{x}^i\|_q > 0)$, where $\boldsymbol{x}^i$ denotes the i-th row of $\boldsymbol{X}$ and $I(\cdot)$ is an indicator function. The sparsity constraint, $\|\boldsymbol{X}\|_{q,0} \leq k$, leads the coefficient matrix to be sparse so that each sample can be represented by a linear combination of at most k representatives. The affine constraint $\mathbf{1}^T\boldsymbol{X} = \mathbf{1}^T$ is imposed on $\boldsymbol{X}$ to ensure the selected representative subset to be invariant with respect to global translation of the data.

The optimization problem in Equation (1) is a NP-hard problem because of the existence of the mixed l_0/l_q norm. Thus a standard l_1 relaxation is used to replace the mixed l_0/l_q norm and accordingly the objective function is rewritten as follow.

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}\|_F^2 \quad s.t. \quad \|\boldsymbol{X}\|_{q,1} \leq \tau, \ \mathbf{1}^T\boldsymbol{X} = \mathbf{1}^T, \quad (2)$$

where $\|\boldsymbol{X}\|_{q,1} \triangleq \sum_{i=1}^N \|\boldsymbol{x}^i\|_q$ is the sparsity constraint and τ is a properly chosen parameter. One can obtain a few representatives according to the indices of the nonzero rows of $\boldsymbol{X}$. Those selected representatives can be seen as a condensed view of the original dataset.

III. The Proposed Method

In this section we first point out the problem we hope to address, then we formulate the objective function of the proposed algorithm in detail, and finally present the optimization method.

A. Problem Statement

In this work, we aim to address the problem of finding a few discriminative representatives which have both representational capacity and discriminant ability. Similar to [2], the representational capacity means that each data point can be expressed as a linear combination of a few representatives instead of all the data points. The discriminant ability indicates that the chosen representatives have proper within-class scatter and between-class scatter. Such selected representatives can nicely describe the whole dataset as well as be discriminative enough for the classification task. The problem of discriminative representatives selection is introduced as follows.

Given a set of training samples $\boldsymbol{Y}$, we want to learn a sparse coefficient matrix $\boldsymbol{X}$, from which a representative subset $\boldsymbol{Y}_{rep} = [\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_2}, \ldots, \boldsymbol{y}_{i_k}] \in \mathbb{R}^{m \times k}$ ($\boldsymbol{Y}_{rep} \subseteq \boldsymbol{Y}$) can be chosen. We propose the following objective function of the structure sparsity based discriminative representative selection model:

$$\min_{\boldsymbol{X}} \quad r(\boldsymbol{Y}, \boldsymbol{X}) + \alpha \|\boldsymbol{X}\|_{q,1} + \beta f(\boldsymbol{X}), \quad (3)$$

where $r(\boldsymbol{Y}, \boldsymbol{X})$ is the reconstruction fidelity term; $\|\boldsymbol{X}\|_{q,1}$ is the sparsity constraint; $f(\boldsymbol{X})$ is the discrimination constraint imposed on the coefficient matrix $\boldsymbol{X}$; α and β are scalar parameters. In the next subsection, we will formulate the objective function in more detail.

B. Formulation

The first part of our model aims at finding a representative subset of a given dataset that can appropriately describe the entire dataset. To overcome the limitation of existing works [1, 3], similar to SMRS [2], we define the reconstruction fidelity term $r(\boldsymbol{Y}, \boldsymbol{X})$ as

$$r(\boldsymbol{Y}, \boldsymbol{X}) = \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}\|_F^2 \quad (4)$$

In practice, each data point will be represented by using as few representatives as possible. To achieve this, a mixed l_1/l_2 norm is imposed to constrain the coefficient matrix $\boldsymbol{X}$, thus the first part of our model is formulated as

$$\min_{\boldsymbol{X}} \quad \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}\|_F^2 + \alpha \|\boldsymbol{X}\|_{2,1}, \quad (5)$$

where $\|\boldsymbol{X}\|_{2,1} \triangleq \sum_{i=1}^N \|\boldsymbol{x}^i\|_2$ is the sparsity constraint and α is a scalar parameter that controls the sparsity of $\boldsymbol{X}$.

The second part of the proposed model $f(\boldsymbol{X})$ aims at making the representative subset $\boldsymbol{Y}_{rep}$ discriminative. We can rewrite $\boldsymbol{Y}$ as $\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ \ldots \ , \boldsymbol{Y}_c]$, where $\boldsymbol{Y}_i$ consists of samples belonging to the i^{th} class. Similarly, we can rewrite $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ \cdots, \ \boldsymbol{X}_c]$, where $\boldsymbol{X}_i$ is a sub-matrix consisting of the coding coefficients on $\boldsymbol{Y}_i$. Then the discriminant ability of the selected representatives can be achieved by imposing some discrimination criterion on the labeled coefficient matrix $\boldsymbol{X}$. In our model, we take the widely used Fisher discrimination criterion [8] into consideration, which can be achieved by minimizing the within-class scatter $S_W(\boldsymbol{X})$ and maximizing the between-class scatter $S_B(\boldsymbol{X})$. Definitions of $S_W(\boldsymbol{X})$ and $S_B(\boldsymbol{X})$ are given as follows:

$$S_W(\boldsymbol{X}) = \sum\nolimits_{i=1}^{c} \sum\nolimits_{\boldsymbol{x}_k \in \boldsymbol{X}_i} (\boldsymbol{x}_k - \boldsymbol{m}_i)(\boldsymbol{x}_k - \boldsymbol{m}_i)^T, \quad (6)$$

$$S_B(\boldsymbol{X}) = \sum\nolimits_{i=1}^{c} n_i (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T, \quad (7)$$

where $\boldsymbol{x}_k$ indicates the k-th column of $\boldsymbol{X}$; $\boldsymbol{m}_i$ and $\boldsymbol{m}$ are mean vectors of $\boldsymbol{X}_i$ and $\boldsymbol{X}$ respectively; and n_i denotes the number of samples belonging to the i^{th} class.

Thus the discrimination item $f(\boldsymbol{X})$ can be defined as

$$f(\boldsymbol{X}) = tr(S_W(\boldsymbol{X})) - tr(S_B(\boldsymbol{X})) + \eta \|\boldsymbol{X}\|_F^2, \quad (8)$$

where $\|\boldsymbol{X}\|_F^2$ is an elastic term, and η is a carefully selected parameter to make $f(\boldsymbol{X})$ be convex. Readers can refer to [9] for more details about the convexity of $f(\boldsymbol{X})$. By imposing the Fisher discrimination criterion on the coefficient matrix $\boldsymbol{X}$, such selected representatives, which can be regarded as a condensed view of the whole dataset, will have the discriminant ability.

Incorporating Equations (4) and (8) into Equation (3), we can obtain the final formulation of the proposed model:

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}\|_F^2 + \alpha \|\boldsymbol{X}\|_{2,1} + \\ & \beta \left(tr(S_W(\boldsymbol{X})) - tr(S_B(\boldsymbol{X})) + \eta \|\boldsymbol{X}\|_F^2 \right). \end{aligned} \quad (9)$$

By minimizing the above objective function, we can select a discriminative representative subset of a given dataset. Those selected representatives can not only appropriately describe the entire dataset but also can be expected to achieve a good performance in terms of classification.

C. Optimization

To optimize the objective function in Equation (9), many convex optimization methods can be employed. In this paper, we implement the widely used Proximal Gradient method [11] to minimize the objective function. To apply the Proximal Gradient method to solve the optimization problem of our model, we rewrite the objective function in Equation (9) as

$$\min_{\boldsymbol{X}} \quad Q(\boldsymbol{X}) + \alpha \Omega(\boldsymbol{X}), \quad (10)$$

where $Q(\boldsymbol{X})$ is a smoothly differentiable function and $\Omega(\boldsymbol{X})$ is a non-differentiable function. $Q(\boldsymbol{X})$ and $\Omega(\boldsymbol{X})$ are defined as follows:

$$\begin{aligned} Q(\boldsymbol{X}) = & \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}\|_F^2 + \\ & \beta \left(tr(S_W(\boldsymbol{X})) - tr(S_B(\boldsymbol{X})) + \eta \|\boldsymbol{X}\|_F^2 \right) \\ = & \|\boldsymbol{Y} - \boldsymbol{Y}X\|_F^2 + \beta \sum_{i=1}^{c} \|\boldsymbol{X}_i - \boldsymbol{M}_i\|_F^2 - \\ & \beta \sum_{i=1}^{c} n_i \|\boldsymbol{m}_i - \boldsymbol{m}\|_2^2 + \beta\eta \|\boldsymbol{X}\|_F^2, \end{aligned} \quad (11)$$

$$\Omega(\boldsymbol{X}) = \|\boldsymbol{X}\|_{2,1}, \quad (12)$$

where $\boldsymbol{M}_i$ is the mean vector matrix (by taking n_i mean vectors $\boldsymbol{m}_i$ as its column vectors) of class i. The Proximal Gradient method is used to optimize Equation (10) by iteratively minimizing the following problem

$$\min_{\boldsymbol{X}} \quad \frac{1}{2} \left\| \boldsymbol{X} - \left(\boldsymbol{X}^t - \frac{1}{L} \nabla Q\left(\boldsymbol{X}^t\right) \right) \right\|_F^2 + \frac{\alpha}{L} \|\boldsymbol{X}\|_{2,1}, \quad (13)$$

where $\boldsymbol{X}^t$ is the coefficient matrix at iteration t, $L > 0$ is a parameter which should be an upper bound on the Lipschitz constant of ∇Q and is typically set with a line-search. According to [10], we can obtain the coefficient matrix at iteration $t+1$, denoted by $\boldsymbol{X}^{t+1}$, as

$$\boldsymbol{X}^{t+1}(i,:) = \begin{cases} \frac{\|\boldsymbol{p}_i\| - \frac{\alpha}{L}}{\|\boldsymbol{p}_i\|} \boldsymbol{p}_i, & if \ \frac{\alpha}{L} < \|\boldsymbol{p}_i\| \\ 0, & otherwise \end{cases} \quad (14)$$

where $\boldsymbol{X}^{t+1}(i,:)$ denotes the i^{th} row of $\boldsymbol{X}^{t+1}$, $\boldsymbol{p}_i \in \boldsymbol{P} = [\boldsymbol{p}_1; \boldsymbol{p}_2; \ldots; \boldsymbol{p}_i; \ldots] = \left(\boldsymbol{X}^t - \frac{1}{L}\nabla Q\left(\boldsymbol{X}^t\right)\right)$.

The proposed algorithm is summarized in Algorithm 1. The coefficient matrix $\boldsymbol{X}$ is iteratively updated untill convergence. Since the indices of nonzero rows of $\boldsymbol{X}$ have correspondence with the indices of columns of the original dataset $\boldsymbol{Y}$, a few representatives of data points can be selected according to the nonzero rows of $\boldsymbol{X}$.

Algorithm 1 The SSDRS algorithm

Input:
Original data matrix $\boldsymbol{Y} \in \mathbb{R}^{m \times N}$;
Number of representatives k;

Output:
The representative subset $\boldsymbol{Y}_{rep} = [\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_2}, \ \ldots \ , \boldsymbol{y}_{i_k}] \in \mathbb{R}^{m \times k} \ (k < N)$

Initialize the coefficient matrix $\boldsymbol{X}$ with zeros
while not converge **do**
  Compute the gradient of $Q(\boldsymbol{X})$ which is denoted by $\nabla Q(\boldsymbol{X})$;
  Set the parameter L with a line-search;
  Update the coefficient matrix $\boldsymbol{X}$ by solving $\min_{\boldsymbol{X}} \ \frac{1}{2} \left\| \boldsymbol{X} - \left(\boldsymbol{X}^t - \frac{1}{L}\nabla Q\left(\boldsymbol{X}^t\right)\right) \right\|_F^2 + \frac{\alpha}{L} \|\boldsymbol{X}\|_{2,1}$ with the Proximal Gradient method [10];
end while
Select k representatives $\boldsymbol{Y}_{rep}$ from the dataset corresponding to the nonzero rows of $\boldsymbol{X}$;

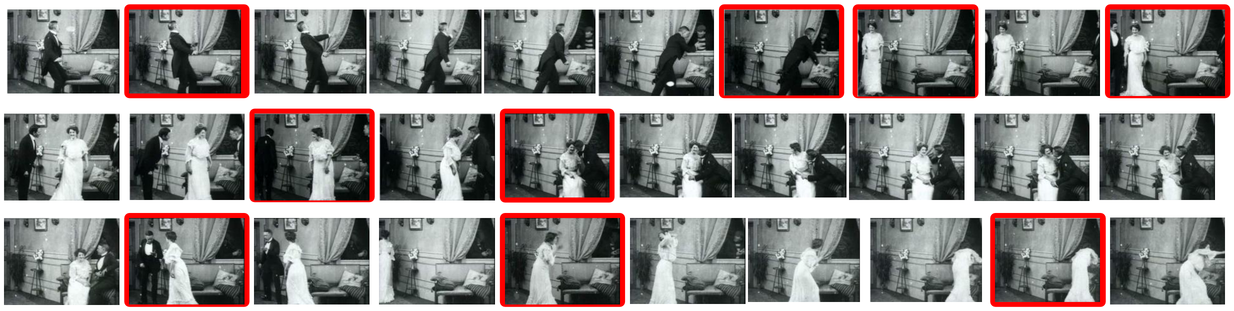

Fig. 1: Brief description of the Society Raffles video and 9 representatives found by our algorithm. Those selected representatives bounded in red rectangles capture main events and well summarize the video. (best viewed in color)

Fig. 2: Some frames of the New Indians multi-shot video and 7 representatives found by our algorithm. Those representatives bounded in red rectangles properly describe the whole frame sequences of the video. Since scene 2 contains more activities than other scenes, our algorithm finds 3 representative frames for scene 2 and 1 representative frame for other scenes. (best viewed in color)

IV. Experimental Results

To evaluate the performance of our proposed algorithm in finding the discriminative representatives of real datasets, we apply the SSDRS algorithm to video summarization and image classification. Since data points in some classes are more difficult to be expressed than those in other categories, we fix the total number of representatives instead of fixing the number of representatives for each class. Furthermore, data points with small pairwise coherence will lead to too-close representatives, thus we prune the representative subset to prevent them from being too close to each other according to [12].

A. Video Summarization

In order to demonstrate the effectiveness of our proposed algorithm, we apply SSDRS to the video summarization task. We firstly choose a 1,536-frame one-shot video from [13]. The video consists of several continuous events under the same background. After employing the SSDRS algorithm, 9 representatives for the video are obtained, which are bounded in red rectangles, as shown in Fig. 1. These 9 representative frames capture main events of the video as follows: a man comes into the room; the man is talking with the thief near the window; a woman appears in the scene; another man gets in accompany with the woman; the man which stands on the left of the woman leaves the room; the woman and the first man sit down and talk with each other; the man steals the woman's tiara and gets ready to leave; the woman sees the thief standing at the window; the woman is fainting on the sofa and the thief has gone. Through Fig. 1, we can see that those representatives selected by our algorithm can well summarize the whole video sequence and capture all different events.

Next, we consider a 706-frame multi-shot video taken from [13]. The video consists of 5 shots, which describe 5 different scenes, and each scene contains continuous activities. We apply SSDRS to the video and obtain 7 representatives. A few frames of the video and the selected representative frames lied in the red rectangles are shown in Fig. 2. Those representative frames summarize the video as follows: a boat is sailing at the sea; a man starts to turn the rudder; the man fixes the rudder in a proper location; the man tries to rotate the rudder back; the other representative frames capture the scene of the boat's bow and sea; the scene of the man's face characteristic and the scene of mountains respectively. From Fig. 2, we can see that those representative frames selected by our algorithm describe the 5 scenes of the video well. Moreover, as scene 2 contains several different continuous activities, 3 representative frames are extracted to describe the continuous activities. Scene 1, scene 3, scene 4, and scene 5 have 1 representative frame respectively.

Experimental results in terms of video summarization demonstrate the effectiveness of our algorithm in finding a representative subset of data points. Those selected representatives can well represent the whole video and are differentiable as well.

B. Image Classification

Furthermore, we evaluate the performance of our selected representatives in the task of image classification.

TABLE I: Classification results on the Extended YaleB database with 7 representatives of 51 training samples in each class.

	NN	NS	SRC	SVM
Rand	30.40%	71.30%	82.60%	87.90%
Kmedoids	37.90%	80.00%	89.10%	94.50%
SMRS	33.80%	84.00%	**93.10%**	**96.80%**
SSDRS	**49.16%**	**85.21%**	92.67%	88.13%

TABLE II: Classification results on the COIL-20 database with 5 representatives of 50 training samples in each class.

	NN	NS	SRC	SVM
Rand	80.00%	81.36%	83.86%	83.18%
Kmedoids	86.36%	86.36%	86.59%	86.82%
SMRS	87.04%	89.77%	91.36%	92.27%
SSDRS	**93.41%**	**93.41%**	**92.05%**	**96.13%**

Datasets: In this paper, all the experiments are conducted on four publicly available datasets. The first dataset is COIL-20 [14], which consists of gray-scale images of 20 objects, and each object contains 72 gray-scale images. The second dataset used in our experiments is the Extended YaleB face database (YaleB) [15]. This database contains face images of 38 individuals under the situation of fixed pose and varying illumination. The third dataset we use is a subset of the USPS digits database (subUSPS) [16], which contains 1000 digits images. This dataset consists of 10 classes corresponding to handwritten digits, and 100 images are randomly selected for each class. The last dataset is the Multiple Feature Data Set (MFEAT) which is taken from UCI Machine Learning Repository [17]. MFEAT is a handwritten database extracted from a Dutch Public utility and contains 2,000 digits images. These images are divided into 10 classes and each class contains 200 digit images.

Settings: In order to demonstrate the effectiveness of our algorithm in finding representatives, we compare our algorithm with simple random selection (Rand) [4], Kmedoids [3] and Sparse Modeling Representative Selection (SMRS) [2]. We employ all the algorithms to select a few representatives, which are treated as the training sets of several standard classification algorithms for performance evaluation. The classifiers include Nearest Neighbor (NN) [8], Nearest Subspace (NS) [10], Sparse Representation-based Classification (SRC) [18] and Linear Support Vector Machine (SVM) [8]. For each class, we randomly select 50 (COIL-20), 51 (YaleB), 50 (subUSPS) and 150 (MFEAT) samples for training and the remaining samples are used for testing. To have a fair comparison for all the algorithms, on average we select 5 (COIL-20), 7 (YaleB), 4 (subUSPS) and 10 (MFEAT) representatives from training samples for each class. As the performance of Kmedoids strongly depends on initialization, we use the same strategy as [2]. For the parameter η used in our algorithm, it is set as 1 for all the experiments. In our method, we fine tune the parameters α and β in Eq. (9) by searching the grid of $\{10^{-5}, 10^{-4}, \ldots 10^{4}, 10^{5}\}$.

Results and Analysis: The classification performance on four datasets is exhibited in TABLE I, TABLE II, TABLE III and TABLE IV respectively. These experimental results indicate that representatives found by our algorithm mostly achieve the highest classification performance on all the datasets.

TABLE III: Classification results on the MFEAT digits database with 10 representatives of 150 training samples in each class.

	NN	NS	SRC	SVM
Rand	75.03%	74.15%	77.20%	76.60%
Kmedoids	78.00%	**78.20%**	77.20%	77.40%
SMRS	77.20%	76.20%	75.20%	79.20%
SSDRS	**78.60%**	76.60%	**78.00%**	**79.20%**

TABLE IV: Classification results on the subUSPS digits database with 4 representatives of 50 training samples in each class.

	NN	NS	SRC	SVM
Rand	57.30%	59.17%	57.61%	65.80%
Kmedoids	62.20%	64.40%	65.80%	69.60%
SMRS	64.60%	69.80%	68.60%	71.00%
SSDRS	**71.40%**	**76.80%**	**72.00%**	**78.60%**

Compared with the state-of-the-art SMRS method, our algorithm has 7% improvement over SMRS averagely for the simplest NN classifier. For NS, SRC and SVM classifiers, our algorithm averagely has 3%, 2% and 1% improvements over SMRS respectively. This indicates that our algorithm works better than SMRS for almost all the classifiers. Moreover, our algorithm achieves significant improvement on COIL-20 and subUSPS, and has some improvements on YaleB and MFEAT in most cases. Compared with SMRS, our algorithm adds the discriminant ability to the selected representatives, thus making SSDRS more suitable for classification tasks. Though a few representatives are selected as the training set, SSDRS achieves good image classification performance based on the representational capacity and discriminant ability of the selected representatives.

Compared with the Rand and Kmedoids methods, our algorithm averagely has 12% and 7% improvements for the simplest NN classifier respectively. We also can see that our algorithm has significant improvements on NS, SRC and SVM classifiers. Moreover, we can see that SMRS outperforms Rand and Kmedoids in most cases. Together with the best classification results our algorithm gains, it indicates that sparse representation based representative selection methods are more effective than the algorithms with cluster structure assumptions. The good performance of Kmedoids in TABLE III may be due to the fact that the MFEAT dataset comes from a low-dimensional subspace, which is the underlying assumption of the NS classifier. Since the Extended YaleB face database consists of images under varying illumination, the inner and intra class distances of training data will be changed when the representative subset is used to replace the entire data. Maybe this is the reason why it leads to the large gap between the NN classifier and other classifiers in TABLE I.

Moreover, we also investigate the classification performance versus different numbers of representatives. To achieve this goal, we implement the proposed algorithm on the COIL-20 dataset to select on average 100, 140, 220 and 300 representatives respectively. In order to have a fair comparison, we select the same number of representatives for the competing meth-

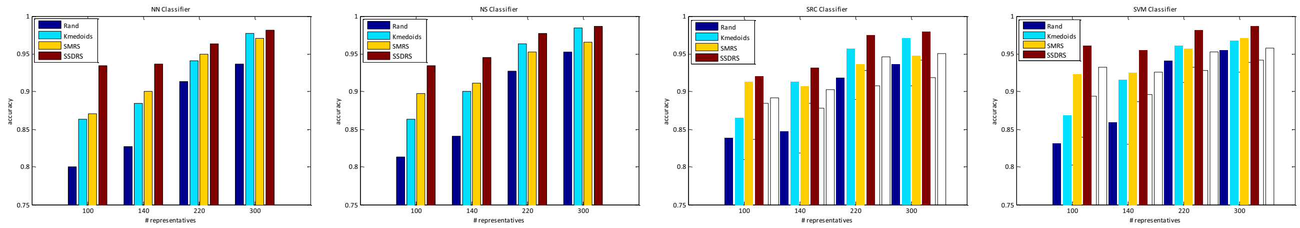

Fig. 3: Accuracies vs. the number of representatives on the COIL-20 database. Classification performances of our algorithm and baselines using NN, NS, SRC and SVM classifiers are exhibited in (a), (b), (c) and (d), respectively.

ods. Experimental results shown in Fig. 3 definitely indicate that our proposed algorithm outperforms the state-of-the-art methods. For all the algorithms, classification performances are improved when the number of representatives increases. This is due to the fact when more representatives are selected, they are more informative about the original dataset. Since our proposed algorithm performs much better than the competing methods when the number of representatives are small, this indicates that the representatives selected by our algorithm are more informative than the competing methods.

V. Conclusion

In this paper, we have proposed a structure sparsity based discriminative representative selection algorithm to find a representative subset of data points which simultaneously have the representational capacity and discriminant ability. Representatives selected by our algorithm not only can appropriately describe the entire dataset but also can lead the representatives belonging to different classes to be discriminative. Furthermore, representatives found by our algorithm can achieve a good classification performance due to the discriminant ability of the representatives. The objective function of our model consists of two parts. The first part is the reconstruction fidelity term which aims at finding a few representatives to express the whole dataset properly; the second part is the discrimination constraint imposed on the coefficient matrix which can ensure those selected representatives to be discriminative. Experiments conducted in terms of video summarization provide an intuitive interpretation for the representation ability and discrimination ability of representatives. And experimental results in terms of image classification also demonstrate the superiority of our proposed algorithm to the state-of-the-art representative selection methods.

Acknowledgment

This work is jointly supported by National Natural Science Foundation of China (61175003, 61135002), Hundred Talents Program of CAS, National Basic Research Program of China (2012CB316300).

References

[1] L. L. Dalton, "Rank revealing qr factorizations," *Linear Algebra and Its Applications*, 1997.

[2] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Computer Vision and Pattern Recognition*, 2012.

[3] L. Kaufman and P. Rousseeuw, "Clustering by means of medoids," *In Y. Dodge(Ed.), Statistical Data Analysis Based on the L1 Norm*, 1987.

[4] J. A. Tropp, "Column subset selection, matrix factorization, and eigenvalue optimization," in *ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[5] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, pp. 972–976, 2007.

[6] L. Balzano, R. Nowak, and W. Bajwa, "Column subset selection with missing data," in *NIPS Workshop on Low-Rank Methods for Large-Scale Machine Learning*, vol. 1, 2010.

[7] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3239–3252, 2012.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, 2012.

[9] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *International Conference on Computer Vision*, 2011.

[10] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *ICML*, 2010.

[11] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[12] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Image Processing*, pp. 4311–4322, 2006.

[13] R. Vidal, "Recursive identification of switched arx systems," *Automatica*, pp. 2274–2287, 2008.

[14] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-20)," *Dept. Comput. Sci., Columbia Univ., New York.[Online] http://www. cs. columbia. edu/CAVE/coil-20. html*, vol. 62, 1996.

[15] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Computer Vision and Pattern Recognition*, vol. 1, 2003.

[16] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 550–554, 1994.

[17] M. Van Breukelen, R. P. Duin, D. M. Tax, and J. d. Hartog, "Handwritten digit recognition by combined classifiers," *Kybernetika*, pp. 381–386, 1998.

[18] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.