SPARSITY PROMOTING DIMENSIONALITY REDUCTION FOR CLASSIFICATION OF HIGH DIMENSIONAL HYPERSPECTRAL IMAGES

Minshan Cui, Saurabh Prasad

University of Houston
Department of Electrical and Computer Engineering
4800 Calhoun Rd. Houston, TX 77004

ABSTRACT

Sparse representation is an active research area in the signal processing and machine learning community in recent years. Recently, sparse representation classifier was proposed for challenging classification tasks — it entails representing a testing sample as a linear combination of all training samples which form an over-complete dictionary. In this paper, we demonstrate that for challenging high-dimensional classification tasks, appropriate dimensionality reduction is beneficial for sparse representation classifiers and it's variants — especially when some features are redundant and/or lack discriminatory power. We propose a new dimensionality reduction algorithm to optimize the performance of greedy pursuit algorithms (required in sparse representation classifiers) by projecting the data into a space where the ratio of intra-class to inter class inner products are maximized. We demonstrate the superiority of the proposed method with standard hyperspectral imagery datasets — both in terms of improved classification accuracy and a speed-up in the run-time.

Index Terms— Sparse Representation, Greedy Pursuit, Dimensionality Reduction

1. INTRODUCTION

Sparse representation is an active research area in the signal processing and machine learning community. In sparse representation, most or all of the information of an unknown signal can be linearly represented by a small number of atoms in a "dictionary". Based on this theory, Wright *et al.* propose a sparse representation classifier (SRC) for robust face recognition [1]. The central idea in SRC is to represent the testing sample as a linear combination of all available training samples (which form an over-complete dictionary) — most of the nonzero or large value entries in the recovered coefficient are expected to correspond to training samples having the same class membership as the testing sample. The assumption of such an approach is that the testing sample approximately lies in the linear span of the training samples from the same class. Experiments conducted in [1] demonstrate that SRC performs well on the face recognition problems under varying lighting conditions.

It is also hypothesized in [1] that the choice of features is not critical when there are sufficient number of features available and sparsity is exploited for the classification task at hand. However, this may not always be the case, particularly when the input feature space possesses very high dimensionality (e.g. with hyperspectral images). Further, some features may be redundant and/or all features may not contain useful discriminative information. Traditional supervised and unsupervised dimensionality reduction algorithms such as Principal Component Analysis (PCA), Locality Preserving Projection (LPP) and Linear Discriminant Analysis (LDA) [2, 3], Regularized LDA etc. may not be optimal for classifiers exploiting sparsity — since the projection in these approaches does not seek to exploit the underlying sparsity in the data.

In recent work, a few dimensionality reduction algorithms aimed at further optimizing the SRC performance have been proposed. Qiao *et al.* proposed a Sparsity Preserving Projection (SPP) [4] method wherein the sparse reconstructive relationship of the samples in the low dimensional subspace are still retained. Since SPP does not utilize class-specific information, it usually lacks of discriminatory power for classification problems. In [5], Lai *et al.* developed a Global Sparse Representation Projections (GSPR) to address this issue. GSPR aims at finding a linearly transformed subspace where global sparse reconstructive relations of the samples are preserved, and interclass samples' separability is also maximized. In an attempt to optimize the performance of SRC, Lan *et al.* and Yang *et al.* present sparse representation based Discriminative Information Exploring Transform (DIET) [6] and SRC steered Discriminative Projection (SRC-DP) [7] respectively. The central idea behind these methods is to project the data onto a space where the ratio of reconstruction errors caused by inter-class samples to intra-class samples are maximized. Since such methods directly maximize the criterion function used in SRC, the performance of SRC is expected to be improved in this projected space. However, the underlying assumption for all of these methods is that the sparse coefficients are recovered accurately. This is due to the fact that the criterion functions of these methods directly use the recovered coefficients to calculate the projection matrix. If the recovered coefficients are inaccurate, the performance of these dimensionality reduction methods can be expected to be unreliable.

In this paper, we propose a new dimensionality reduction method, namely Sparsity Promoting Dimensionality Reduction (SPDR) to optimize the performance of greedy pursuit algorithms such as Orthogonal Matching Pursuit (OMP) [8] and Subspace Pursuit (SP) [9] that are commonly used to recover the sparse coefficients. The key idea in SPDR is to project the data onto a subspace where the ratio of intra-class inner products to inter class inner products are maximized. With such a projection, we expect the accuracy of coefficients recovered by OMP or SP is improved, in a much smaller dimensional space — which also significantly speeds up the coefficient recovery process used in SRC and related algorithms.

Due to its wide application scope, SRC has been gained increased attention in the hyperspectral image analysis community. Owing to rapid advances in sensing technology, hyperspectral imagery is widely available and used for remote sensing, automatic target recognition and surveillance tasks [10, 11]. Hyperspectral im-

E-mail: minshan@ieee.org, saurabh.prasad@ieee.org

agery consists of densely sampled reflectance values over a wide range of the electromagnetic spectrum. The high dimensionality, though potentially useful, poses unique challenges for analysis tasks. Further, for many practical HSI classification tasks, acquiring labeled ground-truth is very expensive and often hard to come-by. Hence, robustness to the over-dimensionality and small-sample-size problem is critical for an effective HSI classification algorithm. In this paper, we validate the proposed method with two standard hyperspectral imagery datasets commonly employed in the hyperspectral image analysis community for benchmarking. We demonstrate the superiority of the proposed method at reducing the dimensionality when followed by some greedy pursuit algorithms by studying the classification accuracy of variants of the SRC algorithm with the proposed feature reduction algorithm.

The rest of this paper is organized as follows. In section 2, we review the SRC classifier. In section 3, we present the proposed SPDR feature reduction algorithm. We provide a description of the experimental classification setup and describe our classification results in section 4, and provide concluding remarks in section 5.

2. SPARSE REPRESENTATION CLASSIFIER

Let $\mathbf{x}_{ij} \in \mathbb{R}^d$ represent the j-th training sample from class i, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_c]$, where $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i}] \in \mathbb{R}^{d \times n_i}$ is the i-th class training sample matrix, c is the number of classes, n_i represents the number of training samples from class i, and n is the total number of training samples, $n = \sum_{i=1}^{c} n_i$. In the context of pattern recognition, our goal is to predict the label of any new testing sample $\mathbf{x} \in \mathbb{R}^d$.

Traditional SRC assumes that a testing sample $\mathbf{x} \in \mathbb{R}^d$ from the i-th class approximately lies in the linear span of training samples from class i. This can be represented as

$$\begin{aligned} \mathbf{x} &\approx \alpha_{i1}\mathbf{x}_{i1} + \alpha_{i2}\mathbf{x}_{i2} + \cdots + \alpha_{in_i}\mathbf{x}_{in_i} \\ &= [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i}][\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in_i}]^T \\ &= \mathbf{X}_i \alpha_i \end{aligned} \tag{1}$$

where α_i is a coefficient vector whose entries are the weights of the corresponding training samples in $\mathbf{X}_i$.

Since the label of the testing sample $\mathbf{x}$ is unknown, $\mathbf{x}$ needs to be represented as a linear combination of all training samples $\mathbf{X}$, which can be formulated as

$$\mathbf{x} = \mathbf{X}\alpha \tag{2}$$

where $\alpha = [\alpha_{11}, \alpha_{12}, \ldots, \alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in_i}, \ldots, \alpha_{cn_c},]$ is a coefficient vector corresponding to all training samples. In an ideal case, if the testing sample $\mathbf{x}$ belongs to i-th class, the entries of α are all zeros except those related to the training samples for the i-th class in $\mathbf{X}$.

After calculating the sparse coefficient α, the residual of each class can be calculated via

$$\mathbf{r}_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{X}_i\hat{\alpha}_i\|_2, \qquad i = 1, 2, \ldots, c \tag{3}$$

where $\hat{\alpha}_i$ denotes the entries of the coefficient vector α associated with the training samples from the i-th class. Finally, $\mathbf{x}$ is assigned a class label i corresponding to a class that resulted in the minimal residual.

Since SRC directly uses the recovered coefficient vector α to determine the membership of the testing sample, accurate recovery of α plays a key role in the efficacy of such a classification algorithm. In practical image processing applications, solving (2) is typically under-determined, one can employ one of many solutions that exist for such a problem. To get the sparsest solution in (2), one can solve the following optimization problem:

$$l_0 : \hat{\alpha} = \operatorname{argmin} \|\alpha\|_0, \quad \text{subject to} \quad \mathbf{X}\alpha = \mathbf{x} \tag{4}$$

where the l_0-norm $\|.\|_0$ counts the number of nonzero entries in α.

In the presence of some noise in the data, the equality constrained problem in (4) can be relaxed into two different inequality constrained problems below:

$$l_0^{'} : \hat{\alpha} = \operatorname{argmin} \|\alpha\|_0, \quad \text{subject to} \quad \|\mathbf{X}\alpha - \mathbf{x}\|_2 \leq \sigma \tag{5}$$

$$l_0^{''} : \hat{\alpha} = \operatorname{argmin} \|\mathbf{X}\alpha - \mathbf{x}\|_2, \quad \text{subject to} \quad \|\alpha\|_0 \leq K \tag{6}$$

where σ is the error tolerance and K is the sparsity level of the solution. Unfortunately, solving these $l_0, l_0^{'}$ and $l_0^{''}$ norm problems are NP-hard and require an exhaustive search through all possible combination of the nonzero entries in α.

As is used in the emerging field of compressive sensing (CS), such an NP-hard problem can be converted to an l_1 norm problem and efficiently solved by a linear programming (LP) known as basis pursuit (BP) [12] or convex programming techniques. It can also be approximately solved by greedy pursuit algorithms such as OMP and SP. The central idea of OMP is to select an atom that produces the highest inner product with the residual of the signal at each iteration and stops when the residual of the signal is lower than some predefined value σ or the number of selected atoms is higher than the predefined sparsity level K. SP iterates in a similar fashion except that it maintains K number of atoms at each iteration instead of one. At the next iteration, K newly selected atoms are added in the current list of atoms and K insignificant atoms are removed from the list. After selecting the atoms, their corresponding coefficients can be calculated by solving an appropriate least squares problem [13].

3. SPARSITY PROMOTING DIMENSIONALITY REDUCTION

In this section, we will describe the proposed Sparsity Promoting Dimensionality Reduction (SPDR) algorithm. Since SRC directly uses the recovered coefficients to predict the membership of the testing sample, accurate recovery of coefficients can be expected to have a siginicant influence on SRC performance. In the proposed SPDR method, we wish to find a lower dimensional subspace via a projection that seeks to optimize the performance of greedy pursuit based recovery methods such as OMP and SP. As we know, OMP and SP select the atoms based on "similarities" (as measured by inner products) with the testing sample. Atoms producing large inner products with the testing sample are likely to be selected. We hence hypothesize that OMP and SP will achieve better recovery performance if between-class samples (belonging to different classes) have large inner products, and within-class samples (belonging to the same class) have small inner products. Therefore, in this paper the criterion function $\mathbf{J}(\mathrm{W})$ for SPDR is chosen to maximize the ratio of within and between class inner products in the projected space which are defined below. Let us assume that $\mathrm{W} \in \mathbb{R}^{\hat{d} \times d}$ is the projection matrix of SPDR where $\hat{d}$ is the reduced dimensionality.

$$\mathbf{J}(\mathrm{W}) = \frac{I_w}{I_b} \tag{7}$$

In (7), I_w and I_b are the within and between class inner products in the projected space, and we define them as below.

$$I_w = \sum_{i=1}^{c}\sum_{j=1}^{n_i}(\mathrm{W}\mathbf{x}_{ij})^T(\mathrm{W}\mu_i) \tag{8}$$

$$I_b = \sum_{i=1}^{c}\sum_{j=1;j\neq i}^{c}(\mathrm{W}\mu_i)^T(\mathrm{W}\mu_j) \tag{9}$$

where $\mu_i = \sum_{j=1}^{n_i}\mathbf{x}_{ij}$ is the mean of the samples in class i.

The optimization process of finding the projection matrix W can be mathematically solved as follows:

$$\begin{aligned}
\hat{\mathrm{W}} &= \underset{W}{\operatorname{argmax}}\,\frac{I_w}{I_b} \\
&= \underset{W}{\operatorname{argmax}}\,\frac{\mathrm{tr}[I_w]}{\mathrm{tr}[I_b]} \\
&= \frac{\mathrm{tr}[(\mathrm{W}\mathbf{x}_{11})^T\mathrm{W}\mu 1 + (\mathrm{W}\mathbf{x}_{12})^T\mathrm{W}\mu 1 + \cdots + (\mathrm{W}\mathbf{x}_{cn_c})^T\mathrm{W}\mu_c]}{\mathrm{tr}[(\mathrm{W}\mu_1)^T\mathrm{W}\mu_2 + (\mathrm{W}\mu_1)^T\mathrm{W}\mu_3 + \cdots + (\mathrm{W}\mu_c)^T\mathrm{W}\mu_{c-1}]} \\
&= \frac{\mathrm{tr}[\mathrm{W}\mu_1(\mathrm{W}\mathbf{x}_{11})^T + \mathrm{W}\mu_1(\mathrm{W}\mathbf{x}_{12})^T + \cdots + \mathrm{W}\mu_c(\mathrm{W}\mathbf{x}_{cn_c})^T]}{\mathrm{tr}[\mathrm{W}\mu_2(\mathrm{W}\mu_1)^T + \mathrm{W}\mu_3(\mathrm{W}\mu_1)^T + \cdots + \mathrm{W}\mu_{c-1}(\mathrm{W}\mu_c)^T]} \\
&= \frac{\mathrm{tr}[\mathrm{W}\mu_1\mathbf{x}_{11}^T\mathrm{W}^T + \mathrm{W}\mu_1\mathbf{x}_{12}^T\mathrm{W}^T + \cdots + \mathrm{W}\mu_c\mathbf{x}_{cn_c}^T\mathrm{W}^T]}{\mathrm{tr}[\mathrm{W}\mu_2\mu_1^T\mathrm{W}^T + \mathrm{W}\mu_3\mu_1^T\mathrm{W}^T + \cdots + \mathrm{W}\mu_{c-1}\mu_c^T\mathrm{W}^T]} \\
&= \frac{\mathrm{tr}[\mathrm{W}\hat{I}_w\mathrm{W}^T]}{\mathrm{tr}[\mathrm{W}\hat{I}_b\mathrm{W}^T]}
\end{aligned} \tag{10}$$

where $\hat{I}_w$ and $\hat{I}_b$ are the within and between class outer products in the original space, defined as:

$$\hat{I}_w = \sum_{i=1}^{c}\sum_{j=1}^{n_i}\mu_i(\mathbf{x}_{ij})^T \tag{11}$$

$$\hat{I}_b = \sum_{i=1}^{c}\sum_{j=1;j\neq i}^{c}\mu_j(\mu_i)^T \tag{12}$$

However this trace-ratio problem does not have a closed form solution. We instead employ and solve for the corresponding ratio-trace problem. Some other approaches to solve this trace-ratio problem can be found in [14, 15]. The conversion of the trace-ratio problem into a ratio-trace problem can be formulated as follows:

$$\frac{\mathrm{tr}[\mathrm{W}\hat{I}_w\mathrm{W}^T]}{\mathrm{tr}[\mathrm{W}\hat{I}_b\mathrm{W}^T]} = \mathrm{tr}[(\mathrm{W}\hat{I}_w\mathrm{W}^T)(\mathrm{W}\hat{I}_b\mathrm{W}^T)^{-1}] \tag{13}$$

Hence the problem of finding the projections that maximizes the criterion function in (7) now reduces to a generalized eigenvalue problem of $\Lambda\hat{I}_b\mathrm{W} = \hat{I}_w\mathrm{W}$, where the projection $\hat{\mathrm{W}}$ are the eigenvectors corresponding to the largest eigenvalues of diagonal matrix of Λ whose diagonal elements are the eigenvalues.

4. EXPERIMENTAL SETUP AND RESULTS

In this section, we validate the efficacy of the proposed SPDR algorithm on two benchmarking hyperspectral datasets. The performance of SPDR is compared with baseline algorithms including LPP, LDA, SPP, DIET and Standard SRC (S-SRC) (where no dimensionality reduction is performed) — in all algorithms, the backend classifier employed was SRC. The performance of these algorithms are assessed by the overall accuracies as a function of training sample size, feature size and their computational time. The reconstructions methods adopted in SRC are based on OMP and SP in an attempt to show that SPDR enhances the performance of these two methods. All free parameters used in SPDR as well as the baseline algorithms are determined by maximizing the classification accuracies by performing an exhaustive grid search over the parameter space.

4.1. University of Pavia Data

The first experimental HSI dataset employed was collected using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor [16]. This image, covering the University of Pavia, Italy, has 103 spectral bands with a spatial coverage of 610×340 pixels, and 9 classes of interests are considered in this dataset. A three-band false color image and its ground-truth are shown in Figure 1. The mean spectral signatures of the nine classes in this dataset are plotted in Figure 2.

(a) (b)

Fig. 1. (a) False color image and (b) Ground-truth of the University of Pavia Data

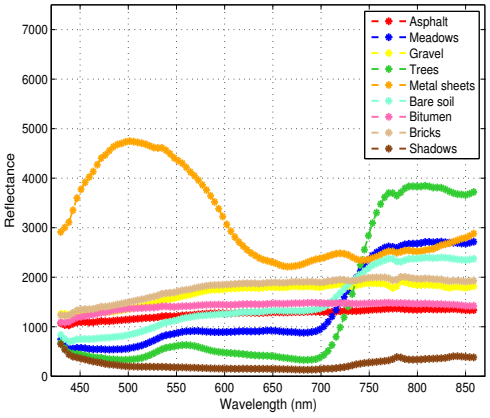

Fig. 2. Mean spectral signatures of University of Pavia Data

The classification accuracies versus different number of training samples for this dataset are reported in Table 1 and Table 2 using OMP and SP as the coefficient recovery methods respectively.

Although it is a relatively simple urban classification problem, the accuracies of SPDR are still higher than other baseline algorithms. Table 3 describes the classification accuracies and testing times in μs. in parenthesis obtained as a function of different feature size. Since the number of features used in LDA is $c-1$ and S-SRC utilizes all of the original spectral features, only one accuracy is reported for each of these two algorithms. It can be seen from this table that the classification accuracies of SPDR are consistently higher than other baseline algorithms with varying number of dimensionality. Note that the testing time of SPDR is much smaller compared with the S-SRC since no dimensionality reduction is performed on it.

Table 1. Classification accuracies(%) obtained for the University of Pavia Data using OMP as the recovery method.

	Number of training samples per class				
Algorithms	10	20	30	40	50
SPDR	71.57	74.73	77.31	77.87	79.10
S-SRC	70.94	74.26	76.40	77.15	78.57
DIET	71.11	74.52	76.58	77.37	78.83
SPP	70.80	74.42	76.31	77.08	78.74
LPP	11.57	53.02	61.20	64.54	66.94
LDA	47.30	55.20	62.66	65.75	67.28

Table 2. Classification accuracies(%) obtained for the University of Pavia Data using SP as the recovery method.

	Number of training samples per class				
Algorithms	10	20	30	40	50
SPDR	72.45	75.67	77.97	78.61	79.63
S-SRC	71.40	74.93	77.11	77.90	78.60
DIET	71.66	75.13	77.24	78.07	78.94
SPP	71.18	74.95	77.00	77.86	78.75
LPP	11.70	54.77	62.79	66.15	68.64
LDA	47.97	56.20	63.95	67.08	69.20

Table 3. Classification accuracies(%) and testing times (μs) in parenthesis obtained for the University of Pavia Data as a function of feature size.

	Number of features				
Algorithms	10	20	30	40	50
SPDR	78.18 (1.9)	78.37 (1.4)	78.50 (4.4)	78.57 (8.6)	78.57 (13.6)
S-SRC	78.57 (13.7)				
DIET	75.14 (8.0)	78.07 (8.9)	78.47 (10.2)	78.53 (11.9)	78.64 (16.1)
SPP	31.87 (7.6)	39.39 (7.1)	57.62 (9.4)	64.96 (11.0)	67.67 (11.9)
LPP	64.83 (5.0)	52.86 (5.0)	44.02 (7.1)	38.39 (9.9)	35.04 (10.7)
LDA	67.28 (1.3)				

4.2. Indian Pines Data

The other dataset used in this work was acquired using NASA's AVIRIS sensor and was collected over northwest Indiana's Indian Pine test site in June 1992. The image represents a vegetation-classification scenario with 145×145 pixels and 220 spectral bands in the 0.4- to 2.45-μm region of the visible and infrared spectrum with a spatial resolution of 20 m. From the 16 different land-cover classes in the image, 7 classes are discarded due to their insufficient number of training samples. 20 noisy bands are removed in the scene covering the region of water absorption and 200 spectral bands are used in the experiments.

The classification results for this dataset using the algorithms described above are reported in Table 4 and Table 5 with the same experimental settings as the University of Pavia dataset. Since the Indian Pines dataset has a relatively large number of densely sampled spectral bands, one can expect there to be substantial redundancies in it. In such a high dimensional space, we can see that the proposed SPDR outperforms other baseline algorithms especially when the number of training samples is small. Also, the accuracies and computational times versus different feature subspace dimensionality are reported in Table 6. Note that the overall classification accuracy of the SPDR is even higher than S-SRC when the dimensionality is reduced to 10.

Table 4. Classification accuracies(%) obtained for the Indian Pines Data using OMP as the recovery method.

	Number of training samples per class				
Algorithms	10	20	30	40	50
SPDR	57.57	62.89	64.44	67.32	68.13
S-SRC	55.32	60.43	62.38	64.81	65.80
DIET	54.92	59.54	61.22	63.29	64.20
SPP	49.46	48.16	53.01	59.87	62.53
LPP	18.30	12.28	39.09	50.38	54.76
LDA	56.84	61.51	63.47	65.49	66.58

Table 5. Classification accuracies(%) obtained for the Indian Pines Data using SP as the recovery method.

	Number of training samples per class				
Algorithms	10	20	30	40	50
SPDR	55.90	62.68	64.41	67.27	68.29
S-SRC	55.50	61.30	63.11	65.28	66.33
DIET	56.09	61.77	63.75	66.10	66.88
SPP	54.38	60.10	61.80	63.76	64.71
LPP	18.15	11.32	43.20	52.43	56.30
LDA	51.17	50.07	54.55	61.34	63.93

Table 6. Classification accuracies(%) and testing times (μs) in parenthesis obtained for the Indian Pines Data as a function of feature size.

	Number of features				
Algorithms	10	20	30	40	50
SPDR	67.66 (7.8)	67.75 (7.6)	67.76 (7.8)	67.69 (7.8)	67.66 (7.8)
S-SRC	65.80 (14.6)				
DIET	65.50 (5.8)	66.27 (6.2)	65.67 (5.6)	65.46 (5.8)	65.52 (5.2)
SPP	28.83 (4.4)	41.55 (6.1)	47.04 (4.8)	55.54 (5.3)	57.74 (5.3)
LPP	51.45 (3.5)	54.27 (3.1)	53.52 (5.0)	52.78 (5.4)	51.77 (6.2)
LDA	62.53 (3.8)				

5. CONCLUSION

In this paper, a new projection based dimensionality reduction method is proposed to optimize the performance of greedy pursuit algorithms in a reduced dimensional space. The proposed algorithm not only enhances the performance of SRC, but significantly reduces the computational cost of invoking OMP or SP — significantly speeding up run-time (testing) with sparse representation classifiers. We expect the proposed approach to provide similar resuts when used in other classification tasks, such as EEG classification, face recognition etc., and we are studying this in ongoing work.

6. REFERENCES

[1] J. A. 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, February 2009.

[2] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing System*, Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, Eds. MIT Press, Cambridge, MA, 2004.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc., New York, 2 edition, 2001.

[4] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, January 2010.

[5] Z. Lai, Z. Jin, and J. Yang, "Global sparse representation projections for feature extraction and classification," in *Chinese Conference on Pattern Recognition*, 2009, pp. 1–5.

[6] C. Lan, X. Jing, S. Li, L. Bian, and Y. Yao, "Exploring the natural discriminative information of sparse representation for feature extraction," in *International Congress on Image and Signal Processing, 2010 3rd*, 2010, vol. 2, pp. 916–920.

[7] J. Yang and D. Chu, "Sparse representation classifier steered discriminative projection," in *International Conference on Pattern Recognition*, 2010, pp. 694–697.

[8] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.

[9] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[10] W. Li, S. Prasad, and J. E. Fowler, "Classification and reconstruction from random projections for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. pp, no. 99, pp. 1 – 11, July 2012.

[11] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1185 – 1198, April 2012.

[12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[13] M. Elad, *Sparse and Redundant Representation: From Theory to Applications in Signal and Image Processing.*, Springer Verlag, 2010.

[14] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 729–735, 2009.

[15] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[16] P. Gamba, "A collection of data for urban area characterization," in *Proceedings of the International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, September 2004, vol. 1, pp. 69–72.