Structured Priors for Sparse-Representation-Based Hyperspectral Image Classification

Xiaoxia Sun, Qing Qu, Nasser M. Nasrabadi, *Fellow, IEEE*, and Trac D. Tran, *Senior Member, IEEE*

***Abstract*—Pixelwise classification, where each pixel is assigned to a predefined class, is one of the most important procedures in hyperspectral image (HSI) analysis. By representing a test pixel as a linear combination of a small subset of labeled pixels, a sparse representation classifier (SRC) gives rather plausible results compared with that of traditional classifiers such as the support vector machine. Recently, by incorporating additional structured sparsity priors, the second-generation SRCs have appeared in the literature and are reported to further improve the performance of HSI. These priors are based on exploiting the spatial dependences between the neighboring pixels, the inherent structure of the dictionary, or both. In this letter, we review and compare several structured priors for sparse-representation-based HSI classification. We also propose a new structured prior called the low-rank (LR) group prior, which can be considered as a modification of the LR prior. Furthermore, we will investigate how different structured priors improve the result for the HSI classification.**

***Index Terms*—Classification, hyperspectral image (HSI), sparse representation, structured priors.**

I. Introduction

One of the most important procedures in hyperspectral image (HSI) is image classification, where the pixels are labeled to one of the classes based on their spectral characteristics. Due to the numerous demands in mineralogy, agriculture, and surveillance, the HSI classification task is developing very rapidly, and a large number of techniques have been proposed to tackle this problem [1]. Compared with previous approaches, support vector machine (SVM) is found highly effective on both computational efficiency and classification results. A wide variety of SVM's modifications have been proposed to improve its performance. Some of them incorporate the contextual information in the classifiers [2], [3]. Others design sparse SVM in order to pursue a sparse decision rule by using ℓ_1-norm as the regularizer [4].

Recently, sparse representation classifier (SRC) has been proposed to solve many computer vision tasks [5], [6], where the use of sparsity as a prior often leads to state-of-the-art performance. SRC has also been applied to HSI classification [7], relying on the observation that hyperspectral pixels belonging to the same class approximately lie in the same low-dimensional subspace. In order to alleviate the problem introduced by the lack of sufficient training data, Haq *et al.* [8] proposed the homotopy-based SRC. Another way to solve the problem of insufficient training data is to employ the contextual information of neighboring pixels in the classifier, such as spectral–spatial constraint classification [9].

In SRC, a test sample $\mathbf{y} \in \mathbf{R}^P$, where P is the number of spectral bands, can be written as a sparse linear combination of all the training pixels (atoms in a dictionary) as

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \tag{1}$$

where $\mathbf{x} \in \mathbf{R}^N$ and $\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$ is ℓ_1-norm. $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$ is a structured dictionary formed from concatenation of several classwise subdictionaries, $\{\mathbf{a}_i\}_{i=1,\ldots,N}$ represents the columns of $\mathbf{A}$, N is the total number of training samples from all the K classes, and λ is a scalar regularization parameter.

The class label for the test pixel $\mathbf{y}$ is determined by the minimum residual between $\mathbf{y}$ and its approximation from each classwise subdictionary

$$\text{class}(\mathbf{y}) = \arg\min_g \|\mathbf{y} - \mathbf{A}\delta_g(\mathbf{x})\|_2^2 \tag{2}$$

where $g \subset \{1, 2, \ldots, K\}$ is the group or class index and $\delta_g(\mathbf{x})$ is the indicator operation zeroing out all elements of $\mathbf{x}$ that do not belong to the class g.

In the case of HSI, SRC always suffers from the nonuniqueness or instability of the sparse coefficients due to the high mutual coherency of the dictionary [10]. Fortunately, a better reconstructed signal and a more robust representation can be obtained by either exploring the dependences of neighboring pixels or exploiting the inherent dictionary structure. Recently, structured priors have been incorporated into HSI classification [7], which can be sorted into three categories: 1) priors that only exploit the correlations and dependences among the neighboring spectral pixels or their sparse coefficient vectors, which include joint sparsity (JS) [12], graph regularized Lasso (referred to as the Laplacian regularized Lasso) [13], and the low-rank (LR) Lasso [14]; 2) priors that only exploit the inherent structure of the dictionary, such as group Lasso [15]; and 3) priors that enforce structural information on both sparse coefficients and dictionary, such as collaborative group Lasso [16] and collaborative hierarchical Lasso (CHiLasso) [17]. Aside from SRC, structured sparsity prior can also be incorporated into other classifiers such as the logistic regression classifiers [18].

The main contributions of this letter are as follows: 1) to assess the SRC performance using various structured sparsity priors for HSI classification and 2) to propose a conceptually

Manuscript received June 3, 2013; revised August 13, 2013 and October 17, 2013; accepted November 7, 2013. Date of publication December 11, 2013; date of current version March 11, 2014. This work was supported in part by the National Science Foundation under Grant CCF-1117545, by the Army Research Office under Grant 60219-MA, and by the Office of Naval Research under Grant N000141210765.

X. Sun, Q. Qu, and T. D. Tran are with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: xsun9@jhu.edu; qqu2@jhu.edu; trac@jhu.edu).

N. M. Nasrabadi is with the U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: nnasraba@arl.army.mil).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2013.2290531

similar prior to CHiLasso, which is called the LR group (LRG) prior. This prior is based on the assumption that pure or mixed pixels from the same classes are highly correlated and can be represented by a combination of sparse LRGs (classes). The proposed prior takes advantage of both the group sparsity (GS) prior, which enforces sparsity across the groups, and the LR prior, which encourages sparsity within the groups, by only using one regularizer.

In the following sections, we investigate the roles of different structured priors imposed on the SRC optimization algorithm. Starting with the classical sparsity ℓ_1-norm prior, we then introduce several different priors with experimental results. The structured priors discussed are JS, Laplacian sparsity (LS), GS, sparse GS (SGS), LR, and LRG priors.

II. HSI Classification via Different Structured Sparse Priors

A. JS Prior

In HSI, pixels within a small neighborhood usually consist of similar materials. Thus, their spectral characteristics are highly correlated. The spatial correlation between neighboring pixels can be indirectly incorporated through a JS model (JSM) [11] by assuming that the underlying sparse vectors associated with these pixels share a common sparsity support. Consider pixels in a small neighborhood of T pixels. Let $\mathbf{Y} \in \mathbf{R}^{P\times T}$ represent a matrix whose columns correspond to pixels in a spatial neighborhood in an HSI. Columns of $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ can be represented as a linear combination of dictionary atoms $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T] \in \mathbf{R}^{N\times T}$ represents a sparse matrix. In JSM, the sparse vectors of T neighboring pixels, which are represented by the T columns of $\mathbf{X}$, share the same support. Therefore, $\mathbf{X}$ is a sparse matrix with only few nonzero rows. The row-sparse matrix X can be recovered by solving the following Lasso problem:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda\|\mathbf{X}\|_{1,2} \tag{3}$$

where $\|\mathbf{X}\|_{1,2} = \sum_{i=1}^{N} \|\mathbf{x}^i\|_2$ is an $\ell_{1,2}$-norm and $\mathbf{x}^i$ represents the ith row of $\mathbf{X}$.

The label for the center pixel $\mathbf{y}_c$ is then determined by the minimum total residual error

$$\text{class}(\mathbf{y}_c) = \arg\min_{g} \|\mathbf{Y} - \mathbf{A}\delta_g(\mathbf{X})\|_F^2 \tag{4}$$

where $\delta_g(\mathbf{X})$ is the indicator operation zeroing out all the elements of $\mathbf{X}$ that do not belong to the class g.

B. LS Prior

In sparse representation, due to the high coherency of the dictionary atoms, the recovered sparse coefficient vectors for multiple neighboring pixels could be partially different even when the neighboring pixels are highly correlated, and this may lead to misclassification. As mentioned in the previous section, JS is able to solve such a problem by enforcing multiple pixels to select exactly the same atoms. However, in many cases, when the neighboring pixels fall on the boundary between several homogeneous regions, the neighboring pixels will belong to several distinct classes (groups) and should use different sets of subdictionary atoms. LS enhances the differences between sparse coefficient vectors of the neighboring pixels that belong to different clusters. We introduce the weighting matrix $\mathbf{W}$, where $\mathbf{w}_{ij}$ characterizes the similarity between a pair of pixels y_i and y_j within a neighborhood. Optimization with an additional LS prior can be expressed as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda_1\|\mathbf{X}\|_1 + \lambda_2 \sum_{i,j} w_{ij}\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \tag{5}$$

where λ_1 and λ_2 are the regularization parameters. The matrix $\mathbf{W}$ is used to characterize the similarity among neighboring pixels in the spectra space. Similar pixels will possess larger weights, therefore enforcing the differences between the corresponding sparse coefficient vectors to become smaller and similarly allowing the difference between sparse coefficient vectors of dissimilar pixels to become larger. Therefore, the LS prior is more flexible than the JS prior in that it does not always force all the neighboring pixels to have the same common support. In this letter, the weighting matrix is computed using the sparse subspace clustering method in [19]. Note that this grouping constraint is enforced on the testing pixels instead of the dictionary atoms, which is different from GS. Let $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ be the normalized symmetric Laplacian matrix [19], where $\mathbf{D}$ is the degree matrix computed from $\mathbf{W}$. We can rewrite (5) as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda_1\|\mathbf{X}\|_1 + \lambda_2 tr(\mathbf{XLX}^T). \tag{6}$$

The above equation can be solved by a modified feature-sign search algorithm [13].

C. GS Prior

The SRC dictionary has an inherent group-structured property since it is composed of several class subdictionaries, i.e., the atoms belonging to the same class are grouped together to form a subdictionary. In sparse representation, we classify pixels by measuring how well the pixels are represented by each subdictionary. Therefore, it would be reasonable to enforce the pixels to be represented by groups of atoms instead of individual ones. This could be accomplished by encouraging coefficients of only certain groups to be active and the remaining groups inactive. Group Lasso [15], for example, uses a sparsity prior that sums up the Euclidean norm of every group coefficients. It will dominate the classification performance particularly when the input pixels are inherently mixed pixels. Group Lasso optimization can be represented as

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \sum_{g\in G} w_g\|\mathbf{x}_g\|_2 \tag{7}$$

where $g \subset \{G_1, G_2, \ldots, G_K\}$, $\sum_{g\in G} \|\mathbf{x}_g\|_2$ represents the group sparse prior defined in terms of K groups, and w_g is the weight and is usually set to the square root of the cardinality of the corresponding group to compensate for the different group sizes. Here, $\mathbf{x}_g$ refers to the coefficients of each group. The aforementioned GS can be easily extended to the case of multiple neighboring pixels by extending problem (7) to collaborative group Lasso, which is formulated as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \sum_{g\in G} w_g\|\mathbf{X}_g\|_2 \tag{8}$$

where $\sum_{g\in G}\|\mathbf{X}_g\|_2$ represents a collaborative group Lasso regularizer defined in terms of group and $\mathbf{X}_g$ refers to each of the group coefficient matrix. When the group size is reduced to one, the group Lasso degenerates into a JS Lasso.

D. SGS Prior

In the formulations (7) and (8), the coefficients within each group are not sparse, and all the atoms in the selected groups could be active. If the subdictionary is overcomplete, then it is necessary to enforce sparsity within each group. To achieve sparsity within the groups, an ℓ_1-norm regularizer can be added to the group Lasso (7), which can be written as

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda_1 \sum_{g\in G} w_g \|\mathbf{x}_g\|_2 + \lambda_2\|\mathbf{x}\|_1. \tag{9}$$

Similarly, (9) can be easily extended to the multiple feature case, which can be written as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2 + \lambda_1 \sum_{g\in G} w_g \|\mathbf{X}_g\|_2 + \lambda_2 \sum_{g\in G} w_g \|\mathbf{X}_g\|_1. \tag{10}$$

Optimization problem (9) is referred to in the literature as the sparse group Lasso and problem (10) as the CHiLasso [17].

E. LR/GS Prior

Based on the fact that spectra of neighboring pixels are highly correlated, it is reasonable to enforce the LR sparsity prior on their coefficient matrix. The LR prior is more flexible when compared with the JS prior which strictly enforces the row sparsity. Therefore, when neighboring pixels are composed of small nonhomogeneous regions, the LR sparsity prior outperforms the JS prior. LR sparse recovery problem has been well studied in [14] and is stated as the following Lasso problem:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_* \tag{11}$$

where $\|\mathbf{X}\|_*$ is the nuclear norm [14].

To incorporate the structure of the dictionary, we now extend the LR prior to group LR prior, where the regularizer is obtained by summing up the rank of every group coefficient matrix

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2 + \lambda \sum_{g\in G} w_g \|\mathbf{X}_g\|_*. \tag{12}$$

The LRG prior is able to obtain the within-group sparsity by minimizing the nuclear norm of each group. Furthermore, the summation of nuclear norms empowers the proposed prior to obtain a GS pattern. Hence, the LRG prior is able to achieve sparsity both within and across groups by using only one regularization term.

III. Results and Discussion

A. Data Sets

We evaluate various structured sparsity priors on two different HSIs and one toy example. The first HSI to be assessed is the Indian Pine, acquired by Airborne Visible/Infrared Imaging Spectrometer, which generates 220 bands, of which 20 noisy bands are removed before classification. The spatial dimension of this image is 145 × 145, which contains 16 ground-truth classes, as shown in Table I. We randomly choose 997 pixels (10.64% of all the labeled pixels) for constructing the dictionary and use the remaining pixels for testing. The second image is the University of Pavia, which is an urban image acquired by the Reflective Optics System Imaging Spectrometer and contains 610 × 340 pixels. It generates 115 spectral bands, of which 12 noisy bands are removed. There are nine ground-truth classes of interests. For this image, we choose 997 pixels (2.32% of all the labeled pixels) for constructing the dictionary and the remaining pixels for testing, as shown in Table II. The toy example consists of two different classes (classes 2 and 14 of the Indian Pine test set), and each class contains 30 pixels. The dictionary is the same as that for the Indian Pine. The toy example is used to evaluate the various sparsity patterns generated by the different structured priors.

TABLE I
Number of Training and Test Samples for the Indian Pine Image

Class	Train	Test
1	6	48
2	137	1297
3	80	754
4	23	211
5	48	449
6	72	675
7	3	23
8	47	442
9	2	18
10	93	875
11	235	2233
12	59	555
13	21	191
14	124	1170
15	37	343
16	10	85
Total	997	9369

TABLE II
Number of Training and Test Samples for the University of Pavia Image

Class	Train	Test
1	139	6713
2	137	1859
3	100	2107
4	133	3303
5	68	1310
6	135	4969
7	95	1261
8	131	3747
9	59	967
Total	997	42926

B. Models and Methods

The tested structured sparse priors are as follows: 1) JS; 2) LS; 3) collaborative GS; 4) SGS; 5) LR; and 6) LRG, corresponding to (3), (6), (8), (10), (11), and (12), respectively. For SRC, the parameters λ, λ_1, and λ_2 of different structured priors range from 10^{-3} to 0.1. Performance on the toy example will be visually examined by the difference between the desired sparsity regions and the recovered ones. For the two HSIs, classification performance is evaluated by the overall accuracy (OA), average accuracy (AA), and the κ coefficient measure on the test set. For each structured prior, we present the result

TABLE III
CLASSIFICATION ACCURACY (IN PERCENT) FOR THE INDIAN PINE IMAGE USING 997 (10.64%) TRAINING SAMPLES

Optimization Techniques		ADMM/SpaRSA							Feature Sign Search	
Class	SVM	ℓ_1	JS	LS	GS	SGS	LR	LRG	ℓ_1	LS
1	77.08	68.75	79.17	85.42	79.17	87.50	75.00	**91.67**	66.67	83.33
2	84.96	58.84	81.94	81.34	80.62	79.92	78.60	81.71	74.42	**89.90**
3	62.67	24.40	56.67	47.35	62.13	76.13	29.87	**89.87**	69.87	78.38
4	8.57	49.52	27.62	49.76	37.14	54.29	15.24	67.62	64.76	**88.15**
5	77.18	81.88	85.46	83.96	84.79	82.55	82.10	83.45	91.72	**94.43**
6	91.82	96.88	98.36	97.48	**98.96**	98.36	98.21	98.36	97.02	98.52
7	13.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**69.57**	0.00
8	96.59	96.59	**100.00**	99.55	99.55	99.55	99.77	99.55	99.55	**100.00**
9	0.00	5.56	0.00	0.00	22.22	0.00	0.00	0.00	**61.11**	0.00
10	71.30	24.00	18.94	31.89	39.95	45.58	8.61	49.60	76.46	**87.43**
11	35.25	96.22	91.63	94.58	91.99	93.02	97.12	92.35	87.62	**98.84**
12	42.39	32.97	45.29	64.68	69.57	65.58	20.83	82.97	78.26	**91.71**
13	91.05	98.95	99.47	99.48	99.47	98.95	98.95	99.47	99.47	**100.00**
14	94.85	98.97	98.97	99.49	98.80	99.31	**99.83**	99.31	97.77	99.57
15	30.70	49.71	55.85	63.84	50.58	80.99	44.15	**89.47**	53.80	69.97
16	27.06	88.24	95.29	97.65	95.29	**98.82**	97.65	97.65	85.88	97.65
OA[%]	64.94	71.17	76.41	79.40	80.19	83.19	71.90	86.46	83.74	**92.58**
AA[%]	56.53	60.72	68.53	64.67	69.39	72.53	59.14	76.43	79.62	**79.87**
κ	0.647	0.695	0.737	0.712	0.781	0.807	0.695	0.843	0.833	**0.923**

TABLE IV
CLASSIFICATION ACCURACY (IN PERCENT) FOR THE UNIVERSITY OF PAVIA IMAGE USING 997 (2.32%) TRAINING SAMPLES

Optimization Techniques		ADMM/SpaRSA							Feature Sign Search	
Class	SVM	ℓ_1	JS	LS	GS	SGS	LR	LRG	ℓ_1	LS
1	84.55	57.11	77.04	95.08	94.01	97.90	91.16	94.15	72.14	**95.85**
2	**82.45**	58.22	67.98	66.70	70.04	68.04	69.73	69.32	59.62	64.28
3	77.08	57.33	44.32	77.55	79.45	73.56	75.80	**79.73**	66.21	76.51
4	94.19	95.94	95.13	95.19	95.31	95.55	95.94	98.46	97.67	**98.97**
5	99.01	**100.00**	99.85	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.85	**100.00**
6	23.55	89.60	88.31	96.60	**100.00**	99.74	**100.00**	99.96	80.60	98.63
7	2.06	83.27	84.38	**96.59**	95.24	95.56	95.06	95.24	86.76	94.69
8	33.89	48,65	65.20	67.36	62.24	44.84	65.24	63.06	75.95	**95.76**
9	53.05	93.69	**99.59**	**99.59**	93.38	93.28	93.57	94.00	90.69	98.35
OA[%]	69.84	66.51	74.05	80.82	81.15	79.07	80.81	81.02	71.41	**81.84**
AA[%]	61.09	75.98	80.06	88.80	87.73	85.36	87.35	87.93	81.05	**91.45**
κ	0.569	0.628	0.681	0.758	0.675	0.624	0.611	0.66	0.672	**0.781**

with the highest OA using cross validation. A linear SVM is implemented for comparison, whose parameters are set in the same fashion as in [7].

In experiments, JS, GS, and LR priors are solved by alternating direction method of multipliers (ADMM) [20], while CHiLasso and Laplacian prior are solved by combining sparse reconstruction by separable approximation (SpaRSA) [21] and ADMM. In addition, in conformity with previous work [13], the Laplacian regularized Lasso is also solved by a modified feature sign search (FSS) method. In this letter, we try to present a fair comparison among all priors. According to the optimization technique, we sort the structured priors into two categories: 1) priors solved by ADMM and SpaRSA and 2) priors solved by FSS-based method. The first rows of Tables III and IV show the methods used to implement the sparse recovery for each structured prior.

C. Results

Sparsity patterns of the toy example are shown in Fig. 1. The expected sparsity regions are shown in Fig. 1(a), where the y-axis labels the dictionary atom index and the x-axis labels the test pixel index. The red and green regions correspond to the ideal locations of the active atoms for classes 2 and 14, respectively. Nonzero coefficients that belong to other classes are shown in blue dots. The JS [Fig. 1(c)] shows clear row sparsity pattern, but many rows are mistakenly activated. As expected, active atoms in Fig. 1(d), (e), and (g) demonstrate GS patterns. Comparing the GS [Fig. 1(d)] and SGS Fig. 1(e), it is observed that most of the atoms are deactivated within groups using SGS.

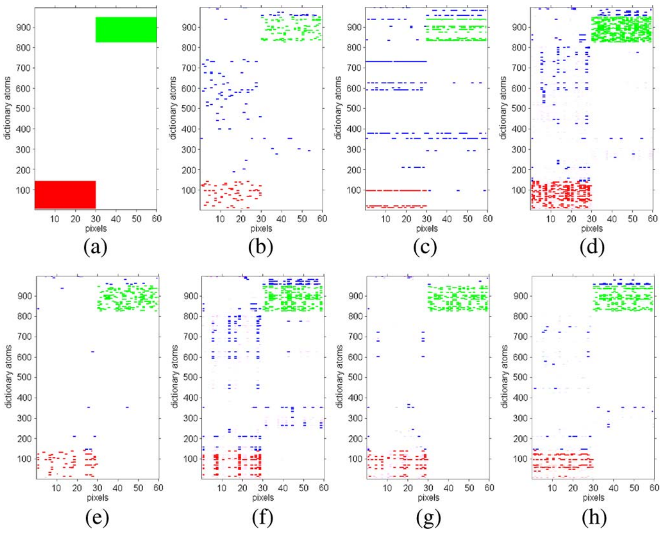

Fig. 1. Sparsity patterns for the toy example: (a) Desired sparsity regions. (b) ℓ_1 minimization using ADMM. (c) JS. (d) Collaborative GS. (e) Collaborative SGS. (f) LR sparsity. (g) LR GS. (h) LS via FFS.

The LRG prior [Fig. 1(g)] demonstrates a similar sparsity pattern as that of SGS. For the LS [Fig. 1(h)], similarity of sparse coefficients that belong to the same classes is clearly visible.

Table III and Fig. 2 show the performance of SRCs with different priors on the Indian Pine image. A spatial window of 9×9 $(T = 81)$ is used since this image consists of mostly large homogeneous regions. Among SRCs with different priors, the worst result occurs when we use simple ℓ_1 ADMM. JS prior gives better result than the LR prior. This is due to the large

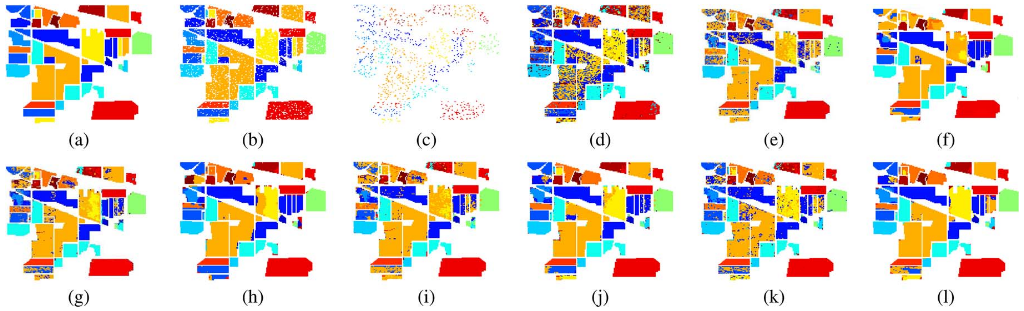

Fig. 2. Results for the Indian Pine image: (a) Ground truth. (b) Training set. (c) Test set. Classification map obtained by (d) SVM. (e) ℓ_1 minimization using ADMM. (f) JS. (g) Collaborative GS. (h) Collaborative SGS. (i) LR sparsity. (j) LR GS. (k) ℓ_1 minimization via FSS. (l) LS via FSS.

TABLE V
COMPUTATION TIME (IN SECONDS) FOR THE INDIAN PINE IMAGE

ADMM/SpaRSA							FFS	
ℓ_1	JS	LS	GS	SGS	LR	LRG	LS	ℓ_1
1124	1874	4015	2811	2649	4403	2904	1124	11628

areas of homogeneous regions in this image, which favors the JSM. The highest OA is given by the LS prior via FFS, and such a high performance is partly contributed to the accurate sparse recovery of the FSS method. Both SGS and LRG outperform GS. We can see that, among ADMM-based methods, the LRG prior yields the smoothest result. The computational time of various structured priors for the Indian Pine image is shown in Table V. Among ADMM/SpaRSA-based methods, LRG, GS, and SGS take roughly similar time (~2500 s) to process the image, while LR and JS require longer time (~4000 s). LS via FFS significantly impedes the computational efficiency.

Results for the University of Pavia image are shown in Table IV. The window size for this image is 5×5 ($T = 25$) since many narrow regions are present in this image. The GS prior gives the highest OA among the priors optimized by ADMM. The LR sparsity prior gives a much better result than JS since this image contains many small homogeneous regions. The LS prior via FFS gives the highest OA performance. However, the difference between performances of various structured priors is quite small.

IV. CONCLUSION

This letter has reviewed five different structured sparse priors and has proposed an LR GS prior. Using these structured priors, classification results of SRCs on HSI are generally improved when compared with those of the classical ℓ_1 sparsity prior. The results have confirmed that the LR prior is a more flexible constraint compared with the JS prior while the latter works better on large homogeneous regions. Imposing the group structured prior on the dictionary always gives higher OA compared with the ℓ_1 prior. We have also observed that the performance not only is determined by the structured priors but also depends on the corresponding optimization techniques.

REFERENCES

[1] A. Plaza, J. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tiltoni, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.

[2] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marì, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[3] L. Gomez-Chova, G. Camps-Valls, J. Muñoz-Marì, and J. Calpe-Maravilla, "Semi-supervised image classification with Laplacian support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 336–340, Jul. 2008.

[4] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani, "1-norm support vector machines," in *Proc. NIPS*, Dec. 2003, vol. 16, pp. 16–23.

[5] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[6] J. Wright, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[7] Y. Chen, M. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[8] Q. Haq, L. Tao, F. Sun, and S. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jun. 2012.

[9] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, and X. Li, "Spectral–spatial constraint hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1811–1824, Mar. 2014.

[10] M. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[11] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[12] E. Berg and M. Friedlander, "Joint-sparse recovery from multiple measurements," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2516–2527, Apr. 2010.

[13] S. Gao, I. Tsang, and L. Chia, "Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, Jan. 2013.

[14] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[15] A. Rakotomamonjy, "Surveying and comparing simultaneous sparse approximation (or group-Lasso) algorithms," *Signal Process.*, vol. 91, no. 7, pp. 1505–1526, Jul. 2011.

[16] S. Kim and E. Xing, "Tree-guided group lasso for multi-task regression with structured sparsity," in *Proc. ICML*, 2010, pp. 543–550.

[17] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. Eldar, "C-HiLasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4183–4198, Sep. 2011.

[18] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[19] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE CVPR*, Jun. 2009, pp. 2790–2797.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2010.

[21] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.