DISCRIMINATIVE SPARSE REPRESENTATIONS IN HYPERSPECTRAL IMAGERY

By

Alexey Castrodad, Zhengming Xing

John Greer, Edward Bosch

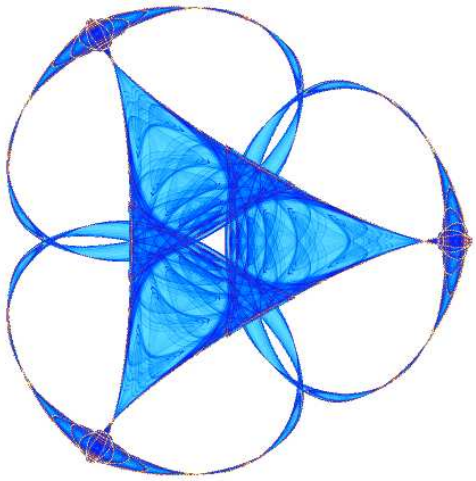
Lawrence Carin

and

Guillermo Sapiro

IMA Preprint Series # 2302

(March 2010)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: http://www.ima.umn.edu

DISCRIMINATIVE SPARSE REPRESENTATIONS IN HYPERSPECTRAL IMAGERY

Alexey Castrodad,[1] *Zhengming Xing,*[2] *John Greer,*[3] *Edward Bosch,*[3] *Lawrence Carin,*[2] *and Guillermo Sapiro*[1]

1. University of Minnesota, 2. Duke University, 3. DoD

ABSTRACT

Recent advances in sparse modeling and dictionary learning for discriminative applications show high potential for numerous classification tasks. In this paper, we show that highly accurate material classification from hyperspectral imagery (HSI) can be obtained with these models, even when the data is reconstructed from a very small percentage of the original image samples. The proposed supervised HSI classification is performed using a measure that accounts for both reconstruction errors and sparsity levels for sparse representations based on class-dependent learned dictionaries. Combining the dictionaries learned for the different materials, a linear mixing model is derived for sub-pixel classification. Results with real hyperspectral data cubes are shown both for urban and non-urban terrain.

Index Terms: Sparse modeling, hyperspectral imagery, classification, dictionary learning.

1. INTRODUCTION

A hyperspectral imager is a powerful tool used for biomedical, environmental, and military applications. HSI is a collection of (possibly hundreds of) narrowly-spaced channels or bands, measuring energy at different wavelengths from the electromagnetic spectrum, and thus allowing spectroscopic analysis. In addition to the geometric spatial features that provide shape information in a typical grayscale or RGB image, HSI also provides spectral features that allow a much richer characterization of the objects and materials in a scene

There are numerous intrinsic challenges associated with HSI. The first one is sensor noise, which is inherent in every electro-optical sensor. There are also complicated light interactions occurring in the atmosphere and on the targeted surface. For example, the atmosphere includes energy from contributing factors such as clouds, haze, and water vapor that need to be corrected. At the surface level, spatial resolution and reflected light off nonuniform surfaces generate spectral mixtures, meaning that the measured energy at each pixel is often not from a homogeneous source but a combination of multiple materials. In addition to these physical factors, the many narrowly-spaced spectral bands yield high-dimensional data, thus making visualization, interpretation, transmission, and exploitation difficult. On the other hand, these spectral bands are highly-correlated and redundant. Consequently there is a need for methods that capitalize on that redundancy to address the processing challenges of high-dimensional data.

Sparse representations express the signal's information with possibly the smallest amount of data from a (usually redundant) dictionary; algorithmically this corresponds to finding a solution to an underdetermined system of linear equations, conditioned/constrained to be sparse (see [1] and references therein). Originally, sparse representations were performed using a fixed dictionary $D \in \Re^{b\times M}$, where M is the number of atoms, and b is the signal's dimension (e.g., DCT, Fourier basis). It is often more appropriate to "learn" these dictionaries and adapt them to the data. State-of-the-art results have been reported in applications related to noise removal, inpainting, discriminative learning, classification, and unsupervised labeling (clustering) [2, 3, 4, 5, 6, 7, 8]. Recently, a non-parametric (Bayesian) approach to sparse modeling and compressed sensing was proposed in [9]. The dictionary is learned using a beta process, which automatically estimates the dictionary size M, and makes no explicit assumption on the noise variance. In addition, it can deal with non-uniform noise sources in the channels, a problem often encountered in HSI. This is the approach used in this paper when reconstructing the HSI from sub-sampled data.

We first propose a framework for supervised full-pixel material identification in remotely sensed HSI using dictionaries that are learned for specific classes. The class label assignment for each pixel is determined by a function that takes into account both the sparsity level and the reconstruction error, and was originally proposed in [7]. Furthermore, we evaluate this technique by validating the data quality of significantly undersampled and then reconstructed HSI following [9]. We address two possible cases. The first case deals with (noisy) training data obtained from the reconstructed image itself. This can be seen as having no *a-priori* information or high quality spectra to match with the spectra in the scene. The second case deals with "high" quality training data, that is acquired from non-subsampled spectra. This could be seen as *a-priori* measurements or knowledge of the contents of the scene or spectra that is acquired in a laboratory. Finally, we deal with spectral mixing by using a combination of atoms from the trained dictionaries.

The remainder of this papers is organized as follows. In Section 2 we describe the proposed approach for HSI supervised classification. Section 3 extends the method to spectral unmixing. Section 4 gives numerical examples, and the last section presents concluding remarks, implications, and future research directions.

2. SUPERVISED CLASSIFICATION OF HSI

In this section, we consider supervised classification. By supervised classification we mean that there are known classes, and for training purposes, known samples pertaining to those classes.

Let the hyperspectral image pixel be represented by the vector valued function $y_i(r,c) : \Re^2 \rightarrow \Re, 1 \leq i \leq b$, where b denotes the number of spectral bands. The following model is assumed during this work: $Y = X + W$, where $Y = [\mathbf{y}_1, ..., \mathbf{y}_n] \in \Re^{b\times n}$ represents the sensor measurements, W is a Gaussian noise source, and X are the "true" signals (target's spectral response). The classification problem becomes that of assigning a label to an estimate of X.

2.1. Learning the HSI dictionaries

Assume there are C possible classes, where C_j is the $j-th$ class representing a pure material. Let the training set for C_j be $\Psi_j =$

$[\psi_1^j, ..., \psi_{n_j}^j]$, a matrix where the column $\psi_i^j \in \Re^b$ is the $i-th$ training sample corresponding to the $j-th$ class. At the training phase of the algorithm, we learn a dictionary (for each class) by solving the following standard sparse modeling problem:

$$(D_j, A_j) := \underset{D,A}{\operatorname{argmin}} \|\Psi_j - DA\|_F^2 + \lambda \|A\|_p, \qquad (1)$$

where $\|\cdot\|_F$ is the matrix Frobenius norm, $D_j \in \Re^{b \times M}$ is the learned dictionary, $A_j = [\alpha_1, ..., \alpha_{n_j}] \in \Re^{M \times n_j}$ is the associated matrix of sparse coefficients, λ is a nonnegative penalty parameter that controls the sparsity of the solution, and p can take the value 0 or 1. When $p = 0$, the l_0 pseudonorm counts the number of nonzero entries in the coefficient vectors. Letting $p = 1$ is a convex relaxation of the problem and is commonly referred to as Lasso [10].[1] The l_1 case tends to be more stable and is preferred during this work.[2] The solution to problem (1) is found using coordinate descent type of algorithms (e.g., KSVD [11]).

2.2. Label assignment

Once the dictionaries are learned, we seek to assign a class label j to each pixel (or block of pixels stacked in column format) to be classified. As proposed in [7], we apply a sparse coding step to the samples $\mathbf{y}$ using each of the learned dictionaries, and simply select the label j corresponding to D_j that gives the minimum value of

$$R(\mathbf{y}, D_j) = \|\mathbf{y} - D_j\alpha\|_2^2 + \lambda\|\alpha\|_1, \quad \forall j, \qquad (2)$$

$\alpha \in \Re^M$. In other words, our classifier is simply the mapping

$$f(\mathbf{y}) = \{j | R(\mathbf{y}, D_j) < R(\mathbf{y}, D_i), i \in [1, ..., C], i \neq j\}. \qquad (3)$$

This means that pixels efficiently represented by the collection of subspaces defined by a common dictionary D_j are classified together. This measure for supervised classification accounts both for reconstruction (fitting) error and sparsity. Without the sparsity term, the classifier (3) can be seen as an *Euclidean Distance Classifier*. The sparsity term especially helps in the presence of noise and/or other artifacts. This naturally comes from the fact that the labeling will tend to prefer the class where the data can be represented in the sparsest way possible, even in cases where the reconstruction error for the tested signal is the same for more than one class. See also [12] for a related penalty when considering the data itself instead of class-dictionaries.

3. SPECTRAL UNMIXING

In the procedure just discussed, there is prior knowledge of the possible sources in the scene, and for each pixel, a label is assigned, corresponding to the class that provides a minimum value in (2). This is a classification at the full-pixel level. It is also possible to extend this to a pixel having one or more labels, implying that it is not composed of a pure class of material, but a combination of these. This is known as *spectral unmixing*, and can be considered as a special case of source separation. The main idea is to decompose each pixel into a linear (or nonlinear) combination of pure sources (i.e., *endmembers*). Focusing in the linear mixing model, a vector with fractional abundances is calculated for each pixel. In an unconstrained case, this can be easily solved using least squares. However, to make the problem physically meaningful, this abundance vector is constrained to be nonnegative and to sum to one, and is known as the *Constrained Least Squares (CLS)* model. It is also desirable that this abundance vector be sparse, meaning that the material at each pixel is explained with as few possible pure sources (see also [12]). A least squares inversion will typically produce a dense solution, however, the sum to one constraint in the CLS model induces a sparse solution. See [13] and references therein for more details in the CLS and other models. More recently, the Least Squares L1 (LSL1) model was proposed for this spectral unmixing problem [12, 14]. In this model, the sum to one constraint was relaxed, meaning an l_1 constraint on the abundance coefficients needs to be minimized, instead of summing strictly one. In addition, as mentioned above, [12] used the data itself instead of learned dictionaries.

An extension to the problem of spectral mixing can be naturally formulated from the framework in Section 2. The model (1) is very similar to what is known as the linear mixing model, where D would represent the materials and A the corresponding abundances. In order to adapt it, we need to add a nonnegativity constraint on both the dictionary and coefficients. Now, compared to the traditional models, where the endmembers are real spectral signatures, here the solution is a linear combination of subspaces representing these endmembers (D are learned atoms and not actual pure materials). One possible advantage of this approach is that it can account for material variability caused for example by factors like noise, non-homogenous substances, etc. The main idea is to train a dictionary for each class, and then form a new dictionary $\overline{D} := [D_1, ..., D_C] \in \Re^{b \times MC}$, similarly in nature to the approach followed in [8] for robust face recognition. In this way, the sparse coding on each pixel comes from a "mixed" union of subspaces (in contrast, [8] expected a single sub-dictionary to be selected at each time). In this work, we use the fully constrained sparse coding step by using a sum to less or equal to one constraint in the abundance coefficients. This is equivalent to solving the sum to one constraint with a zero vector included as an endmember, and therefore allowing shade and dark pixels to be accounted for [15], and addressing the case where there are missing sources. Finally, the problem is solved using a primal-dual strategy. The core algorithm becomes

Input: Hyperspectral scene Y, training sets $\{\Psi_j\}_{j=1}^C$, number of dictionary atoms M, sparsity parameter λ.
Output: Sparse matrix of fractional abundances A for $\mathbf{y}_i, \quad i = 1, ...n$.
Training:

- For each training set $\Psi_j = [\psi_1^j, ..., \psi_{n_j}^j]$, learn $(D_j, A_j) := \operatorname{argmin}_{D \geq 0, A \geq 0} \|\Psi_j - DA\|_F^2 + \lambda\|A\|_1$.
- $\overline{D} := [D_1, ..., D_j], \quad j = 1, ..., C$.

Abundance estimates:

- For each pixel $\mathbf{y}_i$, solve:

$$\alpha_i^* = \arg \min_{\alpha_i \succeq 0, \|\alpha_i\|_1 \leq 1} \|\mathbf{y}_i - \overline{D}\alpha_i\|_2^2.$$

Fig. 1. *Algorithm for sub-pixel supervised classification in HSI.*

[1] The problem in (1) is not convex, however, is biconvex: fixing D makes it convex in A and viceversa.

[2] Experiments were done in this work using both $p = 0$ and $p = 1$. Results using p = 0 are not shown due to space constraints.

4. EXPERIMENTAL RESULTS

A summary and discussion of the experimental results is presented in this section. The first HSI cube tested is the *APHill* scene (with permission from the US Army Engineer Research and Development Center, Topographic Engineering Center, Fort Belvoir, VA), acquired by the HyMAP sensor, with a total of 432,640 pixels. Each pixel is a 106 dimensional vector after removing the high water absorption and noise damaged bands. The second HSI cube tested is the *Urban* scene, acquired by the HyDICE sensor, and has a total of 94,249 pixels, and a subset of 162 channels. It is publicly available at http://www.agc.army.mil/Hypercube/pub/URBAN.zip. The "known" material labels for APHill, and their corresponding training and validation samples are: **C1:** coniferous trees (967, 228); **C2:** deciduous trees (2346, 234); **C3:** grass (1338, 320); **C4:** lake1 (202, 38); **C5:** lake2 (112, 122); **C6:** crop (1026, 58); **C7:** road (197, 50); **C8:** concrete (74, 25); and **C9:** gravel (87, 38). For the Urban scene, the "known" material labels, and the corresponding training samples are: trees (515), grass (289), and road (36).

As mentioned before, there are several objectives for these reported experiments. First, to test the proposed supervised algorithm both at the full-pixel and sub-pixel (spectral unmixing) level. Second, we include results for cases where the data has been reconstructed from significantly subsampled (compressed) images using the technique described in [9]. This assesses how the classification accuracy is degraded when drastically reducing the available measurements. Furthermore, the algorithm is tested under two possible discrimination tasks. The first one makes no *a-priori* knowledge assumption, and attempts to match "known" classes from the scene itself, meaning $\Psi \subset Y$. The second one attempts to match spectra from each class that has been already measured, meaning $\Psi \not\subset Y$. For example, it could be laboratory spectra modified to fit the sensor's characteristics, or previously acquired spectra at full sampling rate (higher quality). For all the experiments, $M = 25$, and $\lambda = 0.01$ and we used the SPAMS software available at http://www.di.ens.fr/willow/SPAMS/.

4.1. Full-pixel labeling

For the first experiment, samples from the image itself are used to train the classifier. Training and validation classification accuracies for each of the 9 classes using the original image, and a reconstruction from only 20% of the original data (with measured pixels and bands selected uniformly at random), are summarized in tables 1 and 2 respectively.[3] Additionally, the accuracies for training and validation sets for several sampling sizes is summarized in Table 5. Pixels with incorrect label assignments most often occurred for the coniferous/deciduous/grass (**C1**/**C2**/**C3**), and road/concrete/gravel (**C7**/**C8**/**C9**) classes. This should not be surprising. First, grass and trees share common spectral features (e.g., high amplitude at the green visible and near infra-red regions). Also, it is common to encounter mixing between those two materials (trees surrounded by grass). Similarly, for the case of concrete and road, spatial resolution plays an important role (sidewalks around roads), but also the fact that concrete and road are spectrally very similar. These effects are increased with the data reconstructed from limited samples, where the spatial interpolation decreases subtle geometric details, and critical spectral resolution may be carried away with the missing data.

[3]The patch dimensions in Table 2 and subsequent tables indicate the size used in [9] for the reconstruction, thereby incorporating spatial coherence in the process. A patch size of $p \times p$ indicates that vectors of dimension bp^2 are used.

C1	C2	C3	C4	C5	C6	C7	C8	C9
0.997	0.990	0.996	1	1	0.998	1	1	1
0.951	1	1	1	1	1	0.72	1	0.973

Table 1. *Per class classification accuracies for the dictionaries learned from the APHill image (without subsampling for this example). First row: classification for training samples. Second row: classification for validation samples.*

C1	C2	C3	C4	C5	C6	C7	C8	C9
0.991	0.972	1	0.985	1	1	0.992	1	1
0.925	1	1	0.973	0.991	1	0.980	0.92	1

Table 2. *Per class classification accuracies for a reconstructed APHill image with 3×3 patches and randomly sampling only 20% of the data. First row: classification for training samples. Second row: classification for validation samples.*

For the second case, where the sources are available *a-priori*, the samples used for the training phase are not extracted from the image to be tested. Instead, the samples were drawn from the original data (fully sampled). This poses a more difficult problem than the first case since the data source is different (needs to be matched to fit the data being tested). This effect can be noticed by looking at Table 3, where the spectral angle, given by $\theta(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2})$, measures how far is the reconstructed data from the original. Fortunately, in this case, the largest angles correspond to the lake1 and lake2 classes. A possible explanation for this is that most of the energy coming from the sun is absorbed by water, and thus the signal to noise ratio is much lower in those regions. Individual classification results for the case of using 20% of the original data are summarized in Table 4.[4] Note that high accuracy is still attained when 80% of the data is missing. In addition, although some of the overall accuracies are low, even when 98% of the data is missing, most of the incorrect labels occurred with classes with strong similarities (e.g., road and concrete). So even very low sampling measurements could provide with relatively accurate, wide-area mappings, as seen in Figure 2 and Table 5.

patch size, data %	Minimum	Maximum	Average	Median
$\mathbf{3 \times 3, 2\%}$	0.5600	65.0973	2.5234	1.8717
$\mathbf{3 \times 3, 5\%}$	0.3119	58.2472	1.4068	1.095
$\mathbf{3 \times 3, 10\%}$	0.2526	23.65	1.0279	0.8505
$\mathbf{3 \times 3, 20\%}$	0.2429	13.2451	0.9275	0.7768
$\mathbf{4 \times 4, 2\%}$	0.4585	77.0751	2.3063	1.6707
$\mathbf{4 \times 4, 5\%}$	0.2917	67.9867	1.45	1.1083
$\mathbf{4 \times 4, 10\%}$	0.2783	19.2132	1.1119	0.9013
$\mathbf{5 \times 5, 2\%}$	0.4099	74.8831	2.2458	1.605

Table 3. *Spectral angle (in degrees) between original and reconstructed sets.*

4.2. Sub-pixel labeling: spectral unmixing

Full-pixel classification provides with a fairly accurate, broad representation of the scene. However, in some cases, and as previously

[4]In Table 4, "training samples" refers to samples in the same spatial location as those used for training in the *a-priori* sources. Due to the sampling and reconstruction process, these samples are not any longer identical to those in the tested image. Same for third column of Table 5.

C1	C2	C3	C4	C5	C6	C7	C8	C9
0.736	0.991	1	0.985	0.991	1	0.746	0.770	0.988
0.442	1	1	0.894	1	1	0.120	0.960	0.973

Table 4. *Per class classification accuracies, using a-priori sources for dictionary learning, for the reconstructed APHill image with* 3×3 *patches and sampling* 20% *of the data. First row: classification for training samples. Second row: classification for validation samples.*

patch size, data %	Training	Validation	Training	Validation
Original	0.9965	0.9851	-	-
$\mathbf{3 \times 3, 2\%}$	0.9561	0.8748	0.7910	0.7514
$\mathbf{3 \times 3, 5\%}$	0.9910	0.9529	0.8745	0.8295
$\mathbf{3 \times 3, 10\%}$	0.9920	0.9845	0.9195	0.8593
$\mathbf{3 \times 3, 20\%}$	0.9898	0.9864	0.9452	0.8903
$\mathbf{4 \times 4, 2\%}$	0.9842	0.9175	0.8288	0.8109
$\mathbf{4 \times 4, 5\%}$	0.9940	0.9727	0.8834	0.8289
$\mathbf{4 \times 4, 10\%}$	0.9951	0.9783	0.9287	0.8617
$\mathbf{5 \times 5, 2\%}$	0.9954	0.9535	0.8412	0.8091

Table 5. *Overall classification accuracies for the original and reconstructed APHill images. The first two columns show overall training and validation results for the case where the training sources are taken from the image. The last two columns show overall training and validation accuracies for the case where fully sampled spectra is available for training.*

suggested, it may be necessary to go further than the full-pixel level in situations where there are sub-pixel targets, or, as commonly encountered in overhead imaging, where partial occlusions may occur due to elevation differences. Consider the example illustrated in Figure 3, for urban HSI. It consists of a road surrounded by grass and trees. A full-pixel detection is unable to give partial information about an occluded (by trees) section of the road (see box in the middle figure), or regions where tree branches and grass are in the same area. Sub-pixel labeling on the other hand provides a clearer characterization of the scene composition, and the variability associated to each of these classes is appropriately accounted with a "learned dictionary endmember."

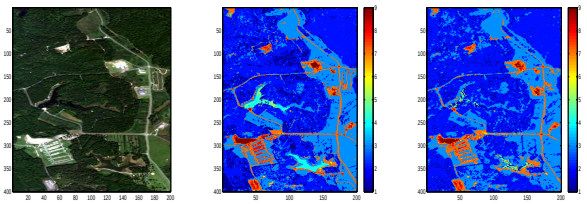


Fig. 2. *Left: False RGB composite of a subset of the APHill scene. Middle: full-pixel classification of original image. Right: full-pixel classification with reconstructed data from* 98% *of the data missing, and* 3×3 *spatial blocks. See [9] for details on the reconstruction technique. (This is a color figure.)*

5. CONCLUDING REMARKS

We proposed a supervised classification algorithm at the full-pixel and sub-pixel levels using learned sparse representations. We reported the results on two hyperspectral datasets, and we also showed the potential for a Bayesian compressed sensing technique to help

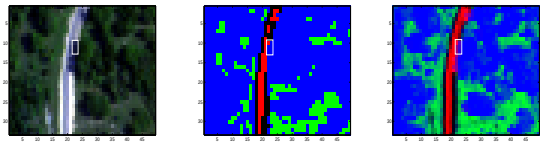


Fig. 3. *Left: False RGB composite of a subset of the Urban scene. The* 3×2 *white rectangle contains a partially occluded section of the road. Middle: Full-pixel classification. The class labels assigned to the pixels inside the rectangle are: road, trees; trees, trees; trees, trees. Note how the road is not complete due to the mixing. Right: Sub-pixel classification. The mixtures obtained in the rectangle are:* 0.07 *trees* + 0.14 *trees* + 0.78 *road,* 0.24 *trees* + 0.46 *trees* + 0.29 *road;* 0.42 *trees* + 0.51 *road,* 0.81 *trees* + 0.19 *road;* 0.46 *trees* + 0.45 *road,* 0.85 *trees* + 0.15 *road. Color is assigned by averaging the nonzero coefficients from each class. (This is a color figure.)*

in solving acquisition, transmission, and storage issues related to HSI. This suggests possible future sensing modes like HSI video and much faster area coverage. Furthermore, noise and data redundancy are managed efficiently by the dictionary learning based classification technique, without the need for explicit dimension reduction or computationally intensive algorithms associated with kernel methods.

6. REFERENCES

[1] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[2] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *CVPR*, 2008, pp. 1–8.

[3] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *NIPS*, 2008, pp. 1033–1040.

[4] J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce, "Discriminative sparse image models for class-specific edge detection and image interpretation," in *ECCV*, 2008.

[5] R. Raina, A. Battle, H Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proceedings of the Twenty-fourth International Conference on Machine Learning*, 2007.

[6] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *ICCV*, 2009.

[7] P. Sprechmann and G. Sapiro, "Dictionary learning and sparse coding for unsupervised clustering," in *ICASSP (to appear)*, 2010.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *PAMI*, vol. 31, no. 2, pp. 210–227, 2008.

[9] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin, "Non-parametric bayesian dictionary learning for sparse image representations," in *NIPS*, 2009.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[12] Z. Guo and S. Osher, "Template matching via l1 minimization and its application to hyperspectral target detection," Tech. Rep. 09-103, UCLA, www.math.ucla.edu/applied/cam/, 2009.

[13] N. Keshava and J.F. Mustard, "Spectral unmixing," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 44–57, 2002.

[14] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its applications to hyperspectral image enhancement," Tech. Rep. 09-30, UCLA, www.math.ucla.edu/applied/cam/, 2009.

[15] M. Velez-Reyes and S. Rosario, "Solving adundance estimation in hyperspectral unmixing as a least distance problem," in *IGARSS*, 2004, vol. 5, pp. 3276–3278.