ROTATION INVARIANCE THROUGH STRUCTURED SPARSITY FOR ROBUST HYPERSPECTRAL IMAGE CLASSIFICATION

Saurabh Prasad, Demetrio Labate, Minshan Cui, Yuhang Zhang

University of Houston

ABSTRACT

Sparse representation based classification has gained popularity with geospatial image analysis in general and hyperspectral image analysis in particular. A central idea with such classification approaches is that a test pixel (spectral reflectance vector) can be sparsely represented in a training dictionary of pixels from all classes – in particular, only training pixels in the dictionary that bear the same class membership of the test pixel will contribute significant coefficients in the sparse representation. The traditional applications of such classifiers to hyperspectral imagery utilize pixel (sample) level information, not spatial contextual information. We propose a sparse representation based classification paradigm that effectively and optimally captures the key geometric properties in hyperspectral images – our classifier that is built on this structured sparse representation then offers very robust classification, including in scenarios where training and test objects have rotational variations (a common occurrence with geospatial images). We validate the proposed approach with benchmark hyperspectral data and present results demonstrating the efficacy of the proposed method.

Index Terms— Hyperspectral, shearlets, morphological separation, structured sparsity.

1. INTRODUCTION

Hyperspectral imaging is becoming increasing popular for a variety of applications such as remote sensing for ground cover analysis, terrestrial/ground based imaging for scene understanding, microscopy and other laboratory imaging for biomedical applications, etc. With the emerging popularity of such multi-channel imagery data, there is also a growing interest in the development of algorithms for robust image analysis. Traditional image analysis with hyperspectral data leverages from the material specific discriminatory information embedded in the spectral reflectance profile of each pixel.

By design, such data contains spectral reflectance over multiple (hundreds to thousands) spectral channels per pixel, while for most applications, the key information content may be in some low dimensional subspace. Feature reduction [1] is hence very useful with such data as a pre-processing to analysis tasks such as classification. Alternately, classification algorithms that exploit the underlying sparsity [2–4] in the data are becoming increasingly popular.

Sparsity has emerged as a powerful tool for a range of applications, including compressed sensing, signal denoising and, more recently, classification. In such representations, 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, a sparse representation classifier (SRC) was developed for robust face recognition, and was later adapted for other applications, including hyperspectral image classification. The central idea in SRC and its variants is to represent a testing sample (e.g. a pixel in a hyperspectral image) as a linear combination of all available training samples (which form an over-complete dictionary) [5–11] — most of the nonzero or large value entries in the recovered coefficients 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. We note that by virtue of their design, such approaches are generally robust to small training sample sizes, even when the dimensionality of the input space is large (e.g. with hyperspectral imagery).

This material is based upon work supported in part by the 2013 NASA New Investigator (Early Career) award (project number: NNX14AI47G) and NSF-DMS (project number: 1320910).

In this work, we present a more holistic approach to sparse representation based classification that leverages from structured sparsity. Specifically, given a hyperspectral image, we build a *morphologically decoupled* sparse representation, resulting in an ensemble of dictionaries, each representing a specific type of information (texture, orientation and scale). We then set up a structured sparse representation based classifier by leveraging from the notion of multi-task joint sparsity over this ensemble of dictionaries, and show that the resulting approach not only captures spatial context very well, but also provides rotation invariance. Rotational invariance allows us to train our classifier on a specific orientation of an object (e.g. a specific orientation of a building), and use that dictionary to effectively classify that same object at other orientations.

This paper is organized as follows. We present the related background and prior work in section 2. We present the proposed approach in section 3, and provide validation with benchmark hyperspectral data in section 4. Concluding remarks are provided in section 5.

2. BACKGROUND AND RELATED WORK

Shearlets emerged during the last decade as a powerful refinement of conventional wavelets and other traditional multiscale representations [12, 13]. Similar to curvelets [14], shearlets are well-localized waveforms defined not only over a range of scales and locations, like wavelets, but also over multiple orientations and with highly anisotropic shapes (see Fig. 1) so that they are especially efficient to capture edges and the other relevant geometric features in images.

In the two (spatial) dimensions, shearlets are generated by the action of anisotropic dilations and shear transformations on a pair of generator functions. By appropriately choosing the generators and adding an appropriate coarse scale system, (see [15] for details), one can define a smooth Parseval frame for $L^2(\mathbb{R}^2)$. Shearlets have been shown to outperform other state of the art multi-scale representation systems such as traditional wavelets, curvelets etc.

Although they have been shown to be very useful for denoising and other related tasks with natural (color) images, their utility for

Fig. 1: **Examples of elements of the shearlet system, at a fixed scale, for different values of the shear parameter.**

classification of high dimensional hyperspectral imagery is relatively unexplored. Another related development with other representations such as wavelets and curvelets has been the notion of morphological separation [16] (partitioning the input image into texture and piecewise smooth components). The framework we propose in this paper involves morphological separation of hyperspectral imagery into texture - a discrete cosine transform (via DCT bases) dictionary, and a cartoon dictionary (via shearlets bases). We then exploit the resulting structured sparsity in the data for robust hyperspectral image classification.

3. PROPOSED APPROACH

In this section, we describe our proposed morphologically decoupled sparse representation (MDSR) approach in detail. To construct our ensemble of dictionaries and take advantage of the properties of shearlets, we adapt an idea originally proposed by us in [17], where we assume that an image x is a superposition of two geometrically distinct components

$$x = x_p + x_t, \tag{1}$$

where x_p is the piecewise smooth component of the data and x_t its textured component. To represent x, we combine two dictionaries $\mathcal{D} = \mathcal{D}_p \cup \mathcal{D}_t$, where $\mathcal{D}_t$ is a local discrete cosine dictionary and is especially efficient for locally periodic patterns; $\mathcal{D}_p$ is a shearlet dictionary and, as recalled above, is optimally sparse for piecewise smooth data. The two dictionaries $\mathcal{D}_p$ and $\mathcal{D}_t$ satisfy the crucial property of being *mutually incoherent* as proved in [18, 19]. That is, each component of x has a sparse representation in one subdictionary but its representation in the other subdictionaries is not sparse. Assuming $x = \mathcal{D}\alpha$, we set the minimization problem:

$$\{\hat{\alpha}_t, \hat{\alpha}_p\} = \min_{\alpha_t,\alpha_p} \eta\left(\|\alpha_t\|_1 + \|\alpha_p\|_1\right) + \frac{1}{2}\left\|x - \mathcal{D}_t\,\alpha_t - \mathcal{D}_p\,\alpha_p\right\|_2^2. \tag{2}$$

Note that, since our subdictionaries are tight frames, $\mathcal{D}_p$ is the Moore-Penrose pseudo inverse of the analysis operator $\mathcal{W}_p$ associated with piecewise smooth data, i.e. $\mathcal{D}_p = \mathcal{W}_p^\dagger$ and, similarly, $\mathcal{D}_t$ is the Moore-Penrose pseudo inverse of the analysis operator $\mathcal{W}_t$ associated with texture data, i.e., $\mathcal{D}_t = \mathcal{W}_t^\dagger$.

Similar to [17], rather than using a sparsity-based *synthesis model* as in (2), we prefer a sparsity-based *analysis model* leading to the minimization problem

$$\{\hat{x}_p, \hat{x}_t\} = \operatorname*{argmin}_{x_p,x_t} \eta\|\mathcal{W}_p x_p\|_1 + \eta\|\mathcal{W}_t x_t\|_1 + \tfrac{1}{2}\|x - x_p - x_t\|_2^2 \tag{3}$$

To further improve the performance, we also include a total variation regularization term which is useful to reducing possible ringing artifacts near the edges introduced by expansion approximations [20]. Thus, we have the optimization problem:

$$\min_{x_p,x_t} \eta\|\mathcal{W}_p\,x_p\|_1 + \eta\|\mathcal{W}_t\,x_t\|_1 + \gamma\,TV(x_p) + \tfrac{1}{2}\|x - x_p - x_t\|_2^2, \tag{4}$$

where TV is the Total Variation, which cane be solved using the iterative shrinkage algorithm introduced by J. Starck et al. [20]. Once the separate estimates $\hat{x}_p$ and $\hat{x}_t$ are obtained as a solution of (4), the final estimator of x is $\hat{x} = \hat{x}_p + \hat{x}_t$. With multi-channel imagery such as HSI, we carry out this separation independently per channel (per individual frame corresponding to each spectral wavelength).

The proposed algorithm is described in Algorithm 1. Here j and ℓ denote the scale and direction indices in the shearlet transform, and j^a and j^f further denote the coarse and fine scales respectively; m denotes the image dimensionality (number of spectral channels) and N_1 represents the number of available training samples. In the first step, the MCA operation is undertaken independently on each spectral channel of the hyperspectral image. This provides two types of dictionaries for SRC based classification: dictionaries corresponding to shearlet coefficients at different scales and orientations representing the cartoon like properties of the image, and dictionaries derived from the recovered DCT image corresponding to texture features. We would like to point out that our use of shearlet analysis coefficients and synthesized texture images is deliberate. By working with shearlet coefficients for classification, our method is able to achieve orientation invariance in classification (in addition to noise robustness) with an appropriate design of the classifier. With regards to the texture component of an image, the synthesized texture image contains image specific texture descriptors as opposed to the raw DCT coefficients which do not carry any information spatially correlated with information in the original image, thereby being unsuitable in the proposed approach.

We use MCA as described previously to build an ensemble of dictionaries — $\mathcal{A}_t$ representing texture components, and $\{\mathcal{A}_p\}^{j\ell}$ representing cartoon components via shearlet coefficients at scale j and orientation ℓ respectively. This sets up our multi-task joint-sparse representation model, where a test sample is simultaneously represented in each of these decoupled components individually, resulting in a weighted global residual over these views. The *min* operation, $\tilde{r}^{l}_{jf} = \min_\ell(\, r^{l}_{jf\ell})$ that we use to minimize residuals across all orientations ℓ at each fine scale j_f is crucial to imparting orientation invariance in the proposed framework. Hence the overall class membership function computes a weighted sum of residuals across the texture and approximation dictionaries, and the minimal residual across orientations at each scale for the fine-scale components. The weighting factors for each dictionary $\{w_{j^a}, w_{jf\ell}, w_t\}$ are estimated as a Fisher's like ratio of between class to within class reconstruction errors

$$\begin{aligned} E_j^{(\mathrm{w})} &= \frac{1}{N_1}\sum_{l=1}^{c}\sum_{i\in\text{class-}l}\|a_i - \mathcal{A}_j\delta_l(\hat{\beta})\|^2, \\ E_j^{(\mathrm{b})} &= \frac{1}{N_1(c-1)}\sum_{l=1}^{c}\sum_{i\in\text{class-}l}\sum_{z\neq l}\|a_i - \mathcal{A}_j\delta_z(\hat{\beta})\|^2, \\ w_j &= \frac{E_j^{(\mathrm{b})}}{E_j^{(\mathrm{w})}}. \end{aligned} \tag{5}$$

where a_i is i-th atom in the dictionary $\mathcal{A}$ and c is the number of classes. These weights scale the residual associated with SRC from each dictionary such that dictionaries that are more discriminative are given preference in the overall decision function.

Algorithm 1 *MDSR*

1: **Input**: A vectorized m-dimensional image $x \in \mathbb{R}^{N^2 \times m}$, test pixel $y \in \mathbb{R}^m$.

{***Morphological Separation***}

2: **for all** $i \in 1, 2, \ldots, m$ **do**

- Calculate the shearlet and DCT coefficients for x^i ($i-$th column of x) based on MCA:

$$\{ \{\hat{\alpha}_p^i\}^{j\ell},\ \hat{\alpha}_t^i \} = \mathit{MCA}\ (\ x^i,\ \mathcal{D}_p,\ \mathcal{D}_t\ ).$$

- Generate the shearlet coefficient matrix for each scale j and each direction ℓ: $\{\mathcal{C}_p^i\}^{j\ell} = \{\hat{\alpha}_p^i\}^{j\ell}$.
- Recover the DCT texture image: $\hat{x}_t^i = \mathcal{D}_t \hat{\alpha}_t^i$.
- Extract texture features from $\hat{x}_t^i$: $\tilde{x}_t^i = \varphi(\hat{x}_t^i)$, where φ denotes a textural feature extractor.

3: **end for**

{***Sparse Representation over Ensemble of Dictionaries***}

4: Assume $\{\mathcal{A}_p \in \mathbb{R}^{N_1 \times m}\}^{j\ell}$ and $\mathcal{A}_t \in \mathbb{R}^{N_1 \times m}$ are the training dictionaries generated from $\{\mathcal{C}_p\}^{j\ell}$ and $\tilde{x}_t$.

5: Obtain representation coefficients ($\{\ \{\hat{\beta}_p\}^{j\ell},\ \hat{\beta}_t\ \}$) corresponding to each dictionary.

{***Morphologically Decoupled Classification***}

6: *Compute residuals:* For the test pixel y for l-th class:

$$r_{j\ell}^l = \|y_{j\ell} - \{\mathcal{A}_p\}^{j\ell} \delta_l(\{\hat{\beta}_p\}^{j\ell})\|_2,$$
$$r_t^l = \|y_t - \mathcal{A}_t \delta_l(\hat{\beta}_t)\|_2.$$

7: *Rotation invariance:* Calculate the minimum residuals of fine scales j^f with regard to different directions d:

$$\tilde{r}_{j^f}^l = \min_{\ell}(\ r_{j^f\ell}^l\ ).$$

8: *Adaptive weighting of residuals:* Use (5) to estimate scaling of residuals corresponding to every dictionary.

9: *Classification:* Determine the class label of a test pixel y based on:

$$\omega = \underset{l=1,2,\ldots,c}{\operatorname{argmin}} \, (w_{j^a} r_{j^a}^l + \sum_{j^f\ell} w_{j^f\ell} \tilde{r}_{j^f\ell}^l + w_t r_t^l).$$

10: **Output:** A class label ω.

To learn the joint sparse representation coefficients in step 5 of the algorithm, the goal is to obtain a row-sparse coefficient matrix which can be modeled as an ℓ_1/ℓ_q-regularized least square problem. For a test sample y^j from source j, given the dictionary $\{\mathcal{A}^j\}_{j=1}^M$ for M sources, the joint sparse coefficient $\mathcal{S} = [\beta^1, \beta^2, ..., \beta^M] \in \mathbb{R}^{n \times M}$ can be estimated by

$$\hat{\mathcal{S}} = \arg\min_{\mathcal{S}} \sum_{j=1}^{M} \left\| y^j - \mathcal{A}^j \beta^j \right\|_2^2 + \lambda \|\mathcal{S}\|_{1,q}, \tag{6}$$

where $\|\mathcal{S}\|_{1,q}$ is the ℓ_1/ℓ_q norm defined as $\|\mathcal{S}\|_{1,q} = \sum_{k=1}^{n} \left\| r^k \right\|_q$, where r^k are the row vectors of $\mathcal{S}$. To make the function convex, q is often set to be greater than 1 (typically 2). Solving the resulting ℓ_1/ℓ_q optimization problem results in a sparse coefficient matrix has common support at the column level. The problem in (6) is convex but non-smooth. An alternating direction method of multipliers (ADMM) [21, 22] is used to solve this optimization problem. Once $\hat{\mathcal{S}}$ is obtained, the class label associated with a test sample is decided by the total minimal residual

$$\omega = \underset{l=1,2,\ldots,c}{\arg\min} \sum_{j=1}^{M} \left\| y^j - \mathcal{A}^j \delta_l(\hat{\beta}^j) \right\|_2^2 \tag{7}$$

where δ_l denotes an indicator function for the l^{th} class — it ensures that only coefficients $\hat{\beta}^j$ that correspond to atoms from the l^{th} class contribute to the residual. Henceforth, we assume that we have c classes in our dictionary and the image. We remark that this approach is particularly suitable to the proposed morphologically decoupled multi-scale framework wherein the image is partitioned into key texture and cartoon components, resulting in M sub-dictionaries $\{\mathcal{A}^j\}_{j=1}^M$ for the hyperspectral image being analyzed. In principle, the proposed framework can utilize (and will be effective for) any sparse representation based classifier at the backend, not just this approach that we chose to validate our framework in this paper.

4. EXPERIMENTAL SETUP AND RESULTS

We validate the proposed approach and compare its efficacy with traditional hyperspectral classification approaches using a real world hyperspectral dataset. The image is acquired using an aerial ITRES-CASI (Compact Airborne Spectrographic Imager) 1500 hyperspectral imager over the University of Houston campus and the neighboring urban area. This geospatial image has spatial dimensions of 1001×281 pixels with a spatial resolution of $2.5m$ per pixel. There are 13 classes and 144 spectral bands over the $380 - 1050nm$ wavelength range, representing common urban classes. Parking lot-1 and Parking lot-2 represent parking lots with and without cars respectively. This dataset was released by us to the research community via the IEEE data fusion contest[1] and covers a wide geographic area over the city of Houston — as a result, it is a challenging dataset with spectral and spatial variability of the various material classes in the scene. It is now an established benchmark dataset in the community [23].

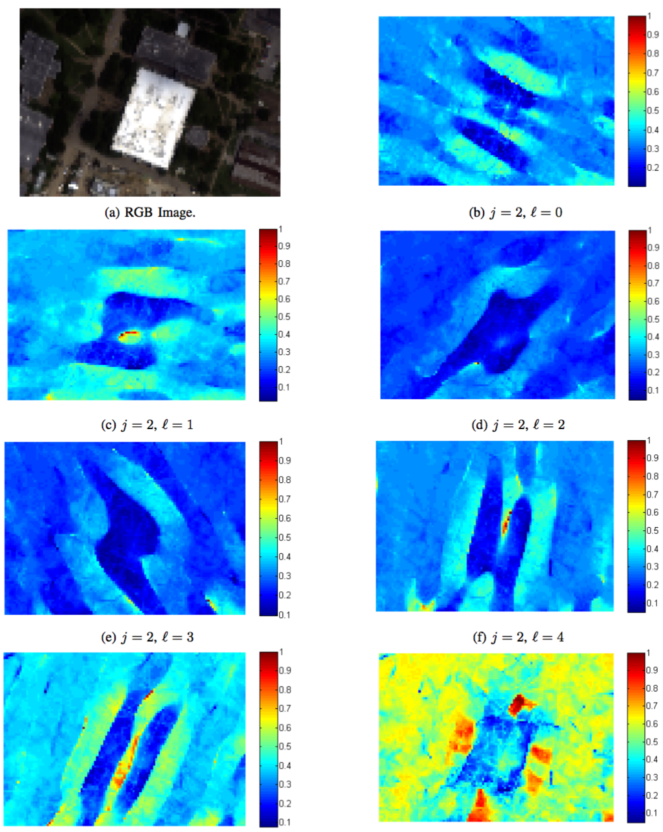
We next summarize key algorithmic parameters used in this paper. We used a two scale of decomposition (each with six orientations) per spectral channel. A Grey Level Co-occurence Matrix (GLCM) based texture feature extractor was utilized for φ in step 2 of Algorithm 1. Specifically, texture features (contrast, entropy, correlation, energy, homogeneity and variance) are extracted over a window around each pixel, with a window size (determined empirically) of 11×11.

Fig. 2 reveals the rotation invariant property imparted by our proposed structured sparse representation based classification — by minimizing the residual over all orientations, we note that in this example (with the "building" class), even when the orientations between the training and the testing pixels is different, the area of minimal residual correctly identifies the structure of the building class. We contend that this is very beneficial in geospatial imaging applications where one often does not have enough exemplars covering all possible rotations of objects — on the contrary, training data is often very limited and ground truth is difficult to come by.

[1] http://hyperspectral.ee.uh.edu/?page_id=459

Table 1: Classification Accuracy as a function of training samples (number of training pixels per class) with the University of Houston (airborne) Hyperspectral Image (average overall accuracies along with standard deviations in parenthesis).

Algorithm / Sample Size		**5**	**10**	**15**	**20**
Proposed (MDSR)	**MD, MS, RI & Weighted**	85.4 (1.6)	91.5 (1.4)	93.4 (1.6)	95.1 (1.2)
	MD, MS & RI	84.5 (1.7)	90.9 (1.7)	93.2 (1.6)	95.1 (1.1)
	MD, MS	77.8 (1.9)	87.6 (2.1)	91.5 (1.5)	93.7 (1.5)
Baseline	**SRC**	81.5 (2.8)	86.3 (1.7)	86.8 (1.1)	86.9 (1.2)



(a) RGB Image. (b) $j = 2, \ell = 0$ (c) $j = 2, \ell = 1$ (d) $j = 2, \ell = 2$ (e) $j = 2, \ell = 3$ (f) $j = 2, \ell = 4$ (g) $j = 2, \ell = 5$ (h) MDSR

Fig. 2: **Residual for the building class for a small cropped portion from the UH dataset, cropped over one of the many buildings in the scene (shown as a natural color image in a), using dictionaries comprised of recovered shearlet coefficients ($r^l_{jf\ell}$) across individual directions (b—g), and using the approach used in MDSR, ($\tilde{r}^l_{jf} = \min_\ell(r^l_{jf\ell})$) — finding the minimum residual across all orientations (h).**

In the next experiment, we report classification accuracy as a function of training sample size (number of training pixels used per class). The number of training pixels was varied from 5 through 20 per class (a commonly studied range for geospatial image analysis [2–4], given the small training sample size often encountered with such applications), while the remaining pixels were used for testing. To further highlight the rotational invariance property and to avoid any bias, we ensured that training and test pixels were not taken from neighboring spatial regions of the aerial imagery. Recall that this being an urban imagery, there are many classes that have a large degree of rotational variability across the scene. These results are summarized in table 1. Specifically, we show results by adding the various components in the proposed framework sequentially (morphological decoupling, multi-scale analysis, rotational invariance and adaptive scaling of residuals). With the proposed framework, we present results with three variations: Morphologically Decoupled, Multi-Scale (*MD, MS*), Morphologically Decoupled, Multi-Scale and Rotational Invariant (*MD, MS & RI*), and Morphologically Decoupled, Multi-Scale, Rotational Invariant and Weighted residuals (*MD, MS, RI, & Weighted*). *MD, MS* connotes a multi-task SRC implementation wherein each scale and orientation of the shearlet coefficients, along with texture features form a dedicated dictionary, and the final classification decision is made by minimizing the sum of residuals over all these dictionaries. In the *MD, MS & RI* approach, instead of accumulating residuals across all orientations and scales, for each scale, we pick the smallest residual over all possible orientations. These "minimum residuals over all orientations" across the various shearlet scales (and texture) are then summed up. In the final variant of the proposed method, *MD, MS, RI, & Weighted*, we weigh individual dictionaries by weights that reflect their relative discriminative ability for the classification task. We compare the performance of this approach to a baseline SRC classifier that is based on spectral features. We note that the proposed approach significantly outperforms the pixel-based variant of SRC. By incorporating rotational invariance, we get a strong performance boost in the accuracy, due to the ability of the classifier to model and exploit spatial context regardless of orientation.

5. CONCLUSIONS

We presented a classification approach that exploits structured sparsity in a sparse representation based classification framework and imparts rotational invariance for classification of high dimensional geospatial imagery (such as hyperspectral remotely sensed imagery). By separating the hyperspectral image cube into sub-dictionaries with specific geometric properties (e.g. piecewise smooth at different scales and orientations, and texture), and proposing an algorithm that imparts rotational invariance, we obtain a significant performance boost in the overall classification performance. We validated the approach with a popular benchmark dataset comprising of an urban hyperspectral scene, and demonstrated that when using the proposed approach to impart rotational invariance and exploit structured sparsity, we get a significant improvement in classification performance.

6. REFERENCES

[1] D Lunga, S Prasad, M Crawford, and O Ersoy, "Manifold-learning-based feature extraction for classification of hyper-

spectral data: A review of advances in manifold learning," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 55–66, 2014.

[2] Minshan Cui and Saurabh Prasad, "Class dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Transactions on Geosciences and Remote Sensing*, vol. 53, no. 5, pp. 2683–2695, May 2015.

[3] Minshan Cui and Saurabh Prasad, "Sparse representation-based classification: Orthogonal least squares or orthogonal matching pursuit?," *Accepted for publication in Pattern Recognition Letters, preprint available at: https://arxiv.org/abs/1607.04942*, 2016.

[4] Minshan Cui and Saurabh Prasad, "Sparsity promoting dimensionality reduction for classification of high dimensional hyperspectral images," in *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013, pp. 2154–2158.

[5] Yi Chen, N.M. Nasrabadi, and T.D. Tran, "Sparse representation for target detection in hyperspectral imagery," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 629–640, 2011.

[6] Chao Lan, Xiao-Yuan Jing, Sheng Li, Lusha Bian, and Yong-Fang Yao, "Exploring the natural discriminative information of sparse representation for feature extraction," in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, 2010, vol. 2, pp. 916–920.

[7] Lei Zhang, Meng Yang, and Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 471–478.

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

[9] A. Wagner, J. Wright, A. Ganesh, Zihan Zhou, and Yi Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 597–604.

[10] Jianchao Yang, J. Wright, T.S. Huang, and Yi Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.

[11] Zihan Zhou, A. Wagner, H. Mobahi, J. Wright, and Yi Ma, "Face recognition with contiguous occlusion using markov random fields," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1050–1057.

[12] G. Kutyniok and D. Labate, *Shearlets: Multiscale Analysis for Multivariate Data*, Springer, 2012.

[13] D. Labate, W. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *SPIE Proc. 5914, SPIE, Bellingham*, pp. 254–262, 2005.

[14] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities.," *Commun. Pure Appl. Anal.*, vol. 57, no. 2, pp. 219–266, 2004.

[15] K. Guo and D. Labate, "The construction of smooth Parseval frames of shearlets.," *Math. Model. Nat. Phenom.*, vol. 8, no. 1, pp. 82–105, 2013.

[16] Michael Elad, J-L Starck, Philippe Querre, and David L Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (mca)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[17] G. R. Easley, D. Labate, and Pooran Negi, "3D data denoising using combined sparse dictionaries.," *Math. Model. Nat. Phenom.*, vol. 8, no. 1, pp. 60–74, 2013.

[18] D. L. Donoho and G. Kutyniok, "Microlocal analysis of the geometric separation problem," *Commun. Pure Appl. Math.*, vol. 66, no. 1, pp. 1–47, 2013.

[19] K. Guo and D. Labate, "Geometric separation of singularities using combined multiscale dictionaries," *J. Fourier Anal. Appl. (in press)*, 2015.

[20] J. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[21] Junfeng Yang and Yin Zhang, "Alternating direction algorithms for \ell_1-problems in compressive sensing," *SIAM journal on scientific computing*, vol. 33, no. 1, pp. 250–278, 2011.

[22] Manya V. Afonso, José M. Bioucas-Dias, and Mário A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 681–695, March 2011.

[23] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Piÿurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2405–2418, June 2014.