Sparse Coding for Spectral Signatures in Hyperspectral Images

Adam Charles
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332–0250
Email: acharles6@gatech.edu

Bruno Olshausen
Helen Wills Neuroscience Institute and School of Optometry
University of California, Berkeley
Berkeley, CA 94720-3220
Email: baolshausen@berkeley.edu

Christopher J. Rozell
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332–0250
Email: crozell@gatech.edu

***Abstract*—The growing use of hyperspectral imagery lead us to seek automated algorithms for extracting useful information about the scene. Recent work in sparse approximation has shown that unsupervised learning techniques can use example data to determine an efficient dictionary with few a priori assumptions. We apply this model to sample hyperspectral data and show that these techniques learn a dictionary that: 1) contains a meaningful spectral decomposition for hyperspectral imagery, 2) admit representations that are useful in determining properties and classifying materials in the scene, and 3) forms local approximations to the nonlinear manifold structure present in the actual data.**

***Index Terms*—E.4: Array Processing and Statistical Signal Processing: Remote Sensing**

I. Introduction

Hyperspectral imagers are airborne sensors that collect ground reflectance measurements across many bands in the electromagnetic spectrum. In hyperspectral imagery (HSI), pixels represent the reflectance at a single ground location in 100+ contiguous spectral bands spanning wavelengths from infrared to ultraviolet. This reflectance data is a function of the spectral signatures of the materials present on the ground, and the high spectral resolution makes HSI a useful modality for determining terrain properties, including material classification, geologic feature identification, and environmental monitoring. HSI is typically analyzed using "endmembers", a library of pure spectral signatures for component materials that form a convex hull containing the measured data (e.g., see the red vectors in Figure 1). Endmember analysis interprets each pixel as a linear combination of endmember vectors, and "unmixes" the material components of each pixel by using a least-squares optimization to determine which endmember vectors best represent the data pixel. In the absence of known endmembers, a number of methods have been proposed to determine endmembers, including expert identification (using spectral libraries or ground truth data), the N-FINDR algorithm (requiring "pure" pixels in the dataset) [1], or other iterative approaches [2].

Recent advances have shown the advantages of using signal models based on sparse decompositions. Sparsity models seek to describe a noisy measurement vector $\mathbf{x}$ using a linear combination of just a few elements from a (perhaps large) dictionary set $\{\phi_k\}$. These types of models have led to state-of-the-art results in many signal and image processing algorithms [3]. Specifically, sparsity models represent a measured signal $\mathbf{x}$ with a noisy linear combination of elements from the dictionary,

$$\mathbf{x} = \sum_{k=1}^{M} \phi_k a_k + \nu, \tag{1}$$

where the coefficients a_k are inferred through minimizing a least-squares cost function with a sparsity-inducing regularization term (e.g., the ℓ_1 norm) [4] such as:

$$J = \left\| \mathbf{x} - \sum_{k=1}^{M} \phi_k a_k \right\|_2^2 + \gamma \sum_k |a_k|. \tag{2}$$

In the model (1) and the cost function (2), ν is additive Gaussian white noise and γ is a scalar parameter trading off between sparsity and approximation accuracy.

Sparsity models appear especially appropriate for analyzing HSI. The high spatial resolution of modern HSI means that it is likely only a few materials will be present in any one pixel, which leads to the belief that each measured pixel should be well-approximated by just a few spectral signatures from a dictionary. Specifically, in the above model the data vector $\mathbf{x}$ (a single hyperspectral pixel) and the dictionary elements ϕ_k would be functions of wavelength (λ). Some initial research into sparsity models for spectral unmixing in HSI has shown promising results [5]. Furthermore, the nature of the sparsity model is to make make the information content of the image explicit by using just a few meaningful dictionary elements for each pixel (as opposed to PCA which represents each pixel with a mixture of basic components), and we believe this representation will be particularly useful for material classification.

With the size of HSI continuing to grow, increased emphasis is being placed on automated methods for analyzing these images. In particular, without ground truth data or assurances that there are "pure" pixels present for every material in the scene, dictionaries to represent the HSI must be learned from the spectrally mixed pixels in the scene. In this work, we show three main results: 1) existing unsupervised learning

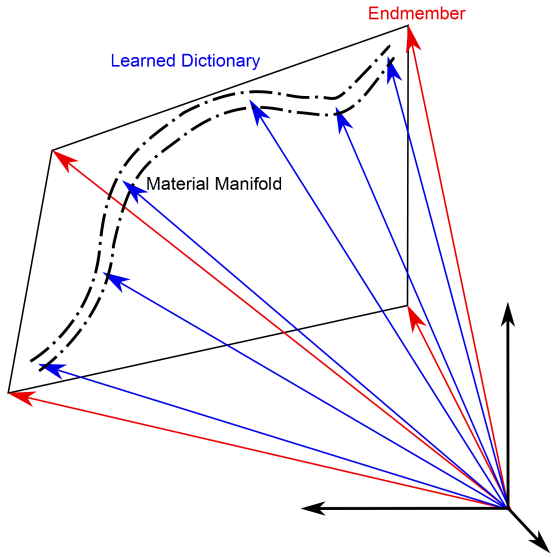

Fig. 1. Endmember analysis uses vectors that compose a convex cone around the data. The red vectors indicate the endmembers for the manifold of data indicated by the two dashed lines. Sparse encoding instead directly uses the spectra along the manifold, indicated here by blue vectors) to linearly approximate the manifold.

algorithms can learn meaningful spectral dictionaries for HSI under a sparsity model, 2) these learned dictionaries admit representations that are useful in determining properties and classifying materials in the scene, and 3) this linear model appears to form local approximations to the nonlinear manifold structure present in the actual data. Preliminary results on similar sparsity-based approaches [6] have shown that sparsity-based models can indeed extract some characteristic material signatures from HSI.

II. Methods

We use the unsupervised learning algorithm first proposed in [7] to learn two sets of dictionaries for the Smith Island HSI (which contains a significant amount of ground truth material labels)[1]. We learned one dictionary with 22 spectral signatures (the number of known material classes in the ground truth labels), and another with 44 spectral signatures. The learning algorithm in [7] attempts to jointly minimize the objective function in Equation (2) with respect to both the coefficients, $\{a_k\}$, and the dictionary, $\{\phi_k\}$. This approach has been successful in learning sparse dictionaries for many types of synthetic and natural signals. For computational tractability, this minimization is often performed alternately on the coefficients and the dictionary elements. At each iteration of the algorithm, we randomly selected a set of 300 pixels and (holding the dictionary constant) found the optimal coefficients $\{a_k\}$ for each pixel[2]. Once the coefficients are found, we adjusted the dictionary $\{\phi_k\}$ by a gradient step of size μ using the update rule

$$\phi_l = \phi_l + \mu \left\langle a_l \left(x - \sum_{k=1}^{M} \phi_k a_k \right) \right\rangle, \tag{3}$$

where $\langle\rangle$ indicates the average over all the randomly selected pixels. Minimizing Equation (2) with respect to $\{\phi_k\}$ can result in a degenerate solution where the norms of the dictionary elements are very large, allowing for near-zero coefficients. To prevent this solution, we restrict each ϕ_k to lie the unit circle. Both dictionaries are learned with $\gamma = 0.02$, and the process is allowed to continue until the dictionary elements are stabilized.

We applied Support Vector Machines (SVMs) to perform supervised classification on the data to demonstrate the utility of the sparse coefficient representations using the learned dictionaries. SVMs were chosen due to their widespread use as a supervised learning algorithm [10]. A total of 2700 pixels from 22 distinct classes were available in the Smith Island dataset. Of these ground truth labels, 70% of the data available for each class is used for training, and the other 30% reserved for testing. We use the multiclass C-SVM classification algorithm implementation available in the `libsvm` package [11] to classify all 22 classes simultaneously. The main parameter of the C-SVM algorithm (C) controls the potential complexity of the classifier by changing the cost of the wrongly classified points in the training process. We performed each classification test over a range of C values to locate value which yield the best possible performance. For each classification test using a different C value in the C-SVM, we used an average of 10 trials to calculate the performance.

III. Results

To test the learned dictionaries, we compare the resulting dictionary elements to known material spectra to see if component materials were in fact learned as predicted by the model. As shown in Figure 2, even with no *a priori* information beyond a sparsity model, the learned spectra examples shown are a good match the spectral signatures of known materials in the image. A number of significantly different spectra are learned, such as "Pine trees", "Water", "Mud" and "Seaoats", as well as very similar spectra, such as "Water" and "Submerged Net" or "Pine trees" and "Iva". Even within classes, small subtleties in the material signatures are captured in the learning. For example, Figure 3 shows that two similar dictionary elements from the 22 element dictionary (shown in the lower half of the image) capture the different characteristics of shallow and deep water. The heat maps in the upper part of the image show the intensity of the corresponding coefficients throughout the image. These heat maps demonstrate that although the two dictionary elements are very similar, the pixels with high magnitudes for the corresponding coefficients are clustered,

[1]This data was taken with a PROBE2 airborne sensor on October 18 2001, and is generously provided to us by the lab of Charles Bachmann. More details on the dataset can be found in [8].

[2]We used the specialized ℓ_1 optimization package `l1_ls` [9] to find the optimal coefficients.

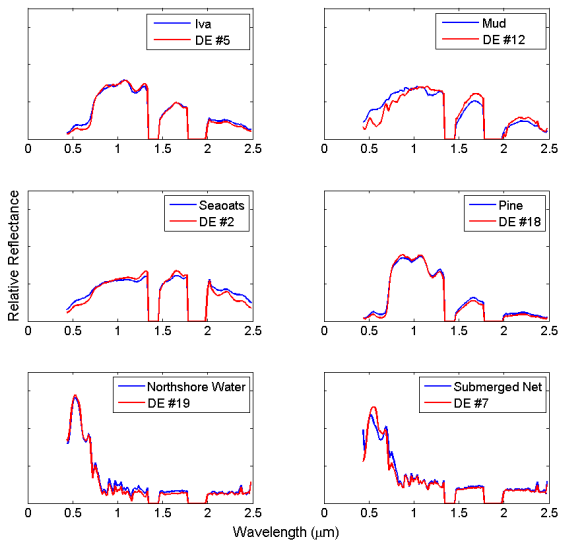

Fig. 2. Example labeled spectra for the materials in the Smith Island data set and the corresponding closest learned dictionary element (DE) for each.

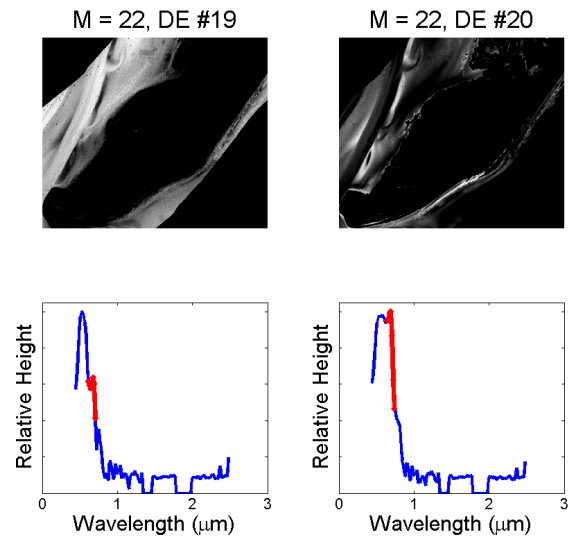

Fig. 3. Subtle differences such as water depth are captured by the learned dictionary elements (DEs), shown in the bottom panels. A low second peak, shown in red, indicates deeper water (left) while a higher second peak indicates shallow areas, such as sandbanks and water closer to shore (right). The top panels show the locations where these dictionary elements are used, corresponding to features such as sandbars.

and more importantly, clustered in different parts of the image. If two coefficients were always active concurrently, they would be indicating the same material and the model would be using two spectra to represent the same material, which conflicts with the model's sparsity assumption. For example, dictionary element #20 in Figure 3 clusters in areas with shallow water, such as around the edge of the island and at sand banks. Dictionary element #19, however, is dominant in most of the deeper water regions, but does not show up by land or in shallow water areas.

Hyperspectral data is nonlinear in nature [12]. Even within a single material class, the image manifold can deviate from a linear model, as shown in Figure 4 for a subset of water pixels from the Smith Island data. It would thus be instructive to compare how the learned dictionaries in this linear model represent these nonlinearities. Figure 4 shows a known nonlinear portion of the data manifold corresponding to bands 14, 29 and 70 (0.6278, 0.8572 and 1.4962 μm) for water pixels [12], and the dictionary elements used to represent that segment of the manifold. The light blue points represent pixels from a portion of the image including a portion of a sand bar, while the multicolored vectors correspond to the dictionary elements from the 44 element dictionary, shown below the scatter plot in Figure 4. The placement of dictionary elements 10, 4, and 26 clearly outline the main manifold curve, while numbers 20 and 22 along with 4 and 10 approximate the cluster of data centered around the end closer to the origin. Thus, although the data actually lies on a nonlinear manifold [12], the learned spectra represent tile this nonlinear space with local approximations.

While the learned spectra can be shown to match known material spectra, a stronger indication of material presence

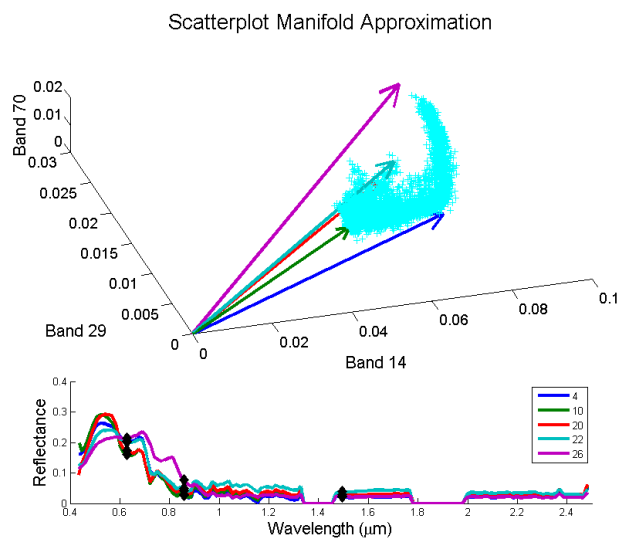

Fig. 4. Manifold structure of water pixels in three spectral bands with the corresponding dictionary elements approximating the manifold (top), along with the full spectra of the dictionary elements in the same color (bottom).

as well as environmental features in the image would be the consistent, meaningful changes in the dictionary coefficients between neighboring pixels. The heat maps in Figure 3 show the presence of various coefficients are fairly consistent (nearby pixels are similarly activated with respect to a specific coefficient). To further test the consistency and to observe the changes over the whole set of coefficients, the decompositions from a row of 21 consecutive pixels in the top row of the Smith Island image were analyzed. The decomposition of

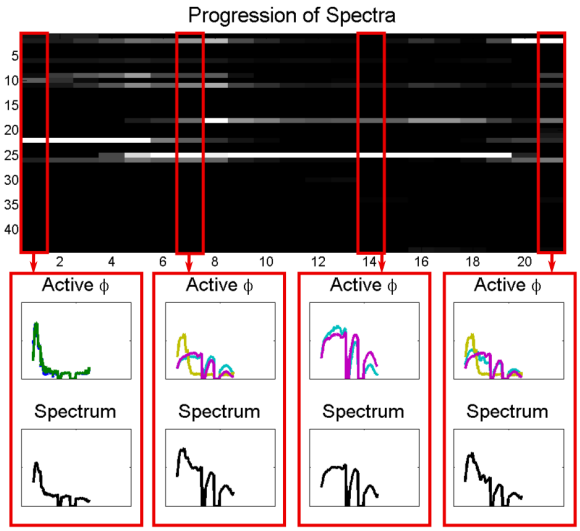

Fig. 5. Progression of the inferred coefficients from water to mud. (Top) The progression through a row of 21 consecutive pixels is shown, with the coefficient numbers on the y-axis, pixel number in the progression on the x-axis, and brightness indicating the level of activation for that coefficient at that pixel. (Bottom) The spectra for every 7th pixel is shown, along with the most active dictionary elements. From pixel 1 to 7 the water depth slowly decreases and turns to mud. From pixel 7 to 14 mud stays consistent and from pixel 14 to 21 the mud gives way to shallow water again.

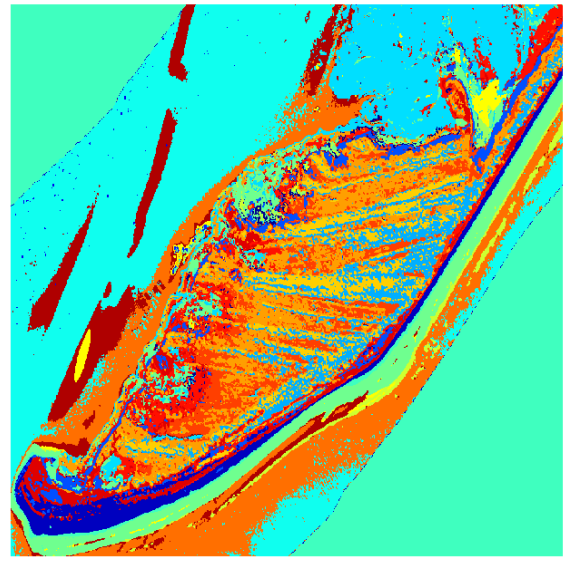

Fig. 6. Vector Quantization classification using only the largest active dictionary element for each pixel. Distinct shapes indicative of known ground truth, such as sandbars and tree lines are clearly visible.

these pixels is shown in Figure 5. This decomposition shows the progression (from left to right), starting at water pixels, progressing to mud and returning back to a mud/water mixture. The consistency of the active coefficients inferred across the progression strongly suggests that dictionary element #25 is representative of a material found in pixels 4-19. Additionally, the consistent transfer both of active coefficients from 4-10 and then again at pixels 18-21 indicate a change in material from one consistently active coefficient to another. To view the behavior of the progression in more detail, the actual spectrum, as well as the 2-3 most prominent dictionary elements in the decomposition are shown for every 7th pixel. The decompositions at the 7th and 21st pixels show the mid-points between two distinct spectral shapes, and the learning algorithm correctly results in a decomposition of the pixels into two distinct spectra rather than one combined spectrum.

For most practical applications, classifying the materials on the ground, either as multiclass material classification or anomaly detection, is one of the prime objectives of HSI. Using the ground truth labels for 22 classes of materials in the Smith Island dataset, we have also explored the utility of the learned dictionaries for material classification. If the learned dictionary spectra are truly indicative of the materials present, then the largest coefficient would be an informative quantity in a classification setting. Figure 6 shows the results of a simple vector quantization scheme where each pixel is classified according to the dictionary element with the largest coefficient. Even this simplistic classification distinctly shows features of the terrain such as locations of sandbanks and tree lines. As can be seen in the center-left portion of the image, a total of three apparent depths of water (illustrated by the colors blue, red and yellow) can be easily identified.

To test the data on a more widely used classification technique, we tested the performance of the inferred coefficients for the learned dictionaries in an SVM classifier (linear kernel) using the set of label pixels from the image. For comparison, we also selected two sets of exemplar data (data chosen from the ground truth to be tested) with comparable sizes (22 and 44 elements), each exemplar dictionary consisting of the same number of spectra from each class (1 and 2 spectra per class respectively). The coefficients for these dictionaries are calculated the same way as the coefficients for the learned dictionaries, using `l1_ls` with $\gamma = 0.02$. Additionally, classification was performed on the raw data as well as the Principal Component Analysis (PCA) decomposition with $M = 4$ principal components (99.9% of the signal energy). Figure 7 shows the results of supervised classification using a linear kernel with C-SVM with the complexity parameter C ranging from 10^{-1} to 10^{4}. We compared two important properties: the classification performance (measured by the percent of spectra misclassified) and the time for classification to be accomplished. The time for the classification represents the complexity of the classifier, and is proportional to the number of support vectors as well as the dimension of the data. As C increases, the classification performance for all representations of the data increases and the number of support vectors, therefore the classification time, decreases. Past large enough values of C, the classification performance and

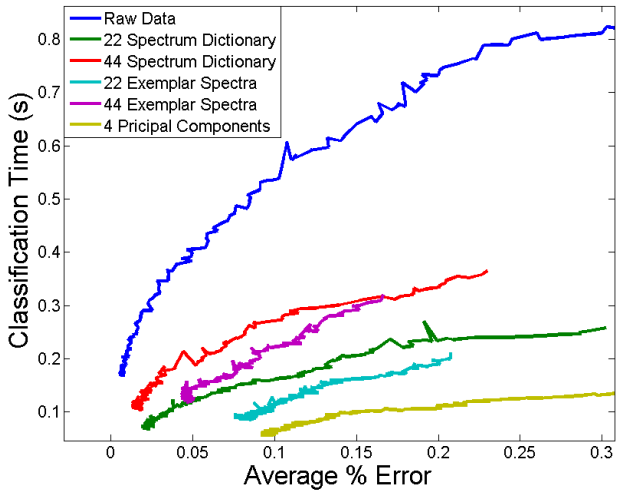

Fig. 7. Support Vector Machine average classification performance using linear kernels, plotting classification error versus classification time. Demixing with either the 22 element dictionary and the 44 element dictionary perform better than demixing with exemplar data (samples drawn from labeled data classes). Performance is comparable with classification on the raw data, but with a lower classification time.

classification complexity plateaued, resulting in the endpoints in Figure 7. While this endpoint was reached at different values of C for each representation of the data, these points are the indicators of the true potential of classification using that representation. Figure 7 shows that classification using the first four principal components is by far the fastest in terms of classification time, but has almost three times the classification error using either of the the learned dictionary representations. Using the exemplar dictionaries yields better classification results than PCA, but also a higher rate of misclassified pixels than the learned dictionaries. The classification time, though, for a given percent error, is less for the exemplar dictionaries as opposed to the learned dictionaries of the same size, indicating that the learned dictionaries use less support vectors. In comparison with classification over the raw data, the learned dictionaries using 44 and 22 elements perform material classification almost as well in terms of percent misclassified pixels. The classifiers, however, are much simpler for the learned dictionaries (the support vectors are lower dimensional), resulting in faster classification times for significantly better classification rates than achievable using either PCA or the exemplar dictionaries.

IV. Conclusions

The results from the learned dictionaries indicate that unsupervised learning of sparse hyperspectral dictionaries can be a useful tool for analyzing HSI. The ability to differentiate, for example, different depths of water, or to locate basic terrain information such as tree lines with simplistic operations could be useful for navigation in areas with no previous ground exploration required. Even in terrain with some ground truth available, the classification performance indicates that materials can be found with higher confidence using the learned dictionary with supervised learning rather than using only the ground truth directly in the same supervised classification setting. Due to the dimensionality reduction Such classification can also be accomplished faster than using the raw data. These aspects of classification using the learned dictionary in combination with the autonomous nature of the learning algorithm makes for an ideal analysis tool for large geological surveys with little or no ground truth.

References

[1] M. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *IEEE Porceedings of SPIE*, vol. 3753, no. 5, 1999.

[2] D. Rogge, B. Rivard, J. Zhang, and J. Feng, "Iterative spectral unmixing for optimizing per-pixel endmember sets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, Dec 2006.

[3] M. Elad, M. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *IEEE Proceedings - Special Issue on Applications of Compressive Sensing & Sparse Representation*, Oct 2008.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[5] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," Feb 2010.

[6] J. Greer, "Sparse demixing," *IEEE Proceedings of SPIE*, vol. 7695, May 2010.

[7] B. A. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 13, pp. 607–609, Jun. 1996.

[8] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2786–2803, 2006.

[9] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large scale l1-regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, Dec 2007.

[10] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

[11] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[12] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploited manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.