# Hyperspectral Band Selection and Endmember Detection using Sparsity Promoting Priors

Alina Zare *Student Member, IEEE*, Paul Gader *Senior Member, IEEE*
Department of Computer Information Science and Engineering, University of Florida, Gainesville, FL

***Abstract*— This paper presents a simultaneous band selection and endmember detection algorithm for hyperspectral imagery. This algorithm is an extension of the Sparsity Promoting Iterated Constrained Endmembers (SPICE) algorithm. The extension adds spectral band weights and a sparsity promoting prior to the SPICE objective function to provide integrated band selection. In addition to solving for endmembers, the number of endmembers, and endmember fractional maps, this algorithm attempts to autonomously perform band selection and determine the number of spectral bands required for a particular scene. Results are presented on a simulated dataset and the AVIRIS Indian Pines dataset. Experiments on the simulated dataset show the ability to find the correct endmembers and abundance values. Experiments on the Indian Pines dataset show strong classification accuracies in comparison to previously published results.**

***Index Terms*—Sparsity Promotion, Endmember, Hyperspectral Imagery, Band Selection, Dimensionality Reduction**

## I. Introduction

The goal of this paper is to perform simultaneous band selection and endmember detection. Performing simultaneous band selection and endmember determination allows endmember information to be used while performing band selection. This is accomplished by extending SPICE [1], an endmember detection algorithm, to incorporate hyperspectral band selection. The proposed algorithm performs band selection and determines the number of bands required for the input hyperspectral scene while simultaneously determining endmembers and the number of endmembers required.

Many data reduction techniques such as Principal Components Analysis (PCA) and Maximum Noise Fraction transform (MNF) [2],[3] have been used to project the data into a lower dimensional space and thus reduce the dimensionality of the data. Although these methods are effective at data reduction, they do not retain physically meaningful bands that correspond to wavelengths in the original dataset. Harsanyi and Chang provide an orthogonal subspace projection approach that also transforms the data [4]. DeBacker, et al. and Kumar, et al. both present methods that merge many adjacent hyperspectral bands losing the one-to-one correlation with the original dataset [5],[6]. Instead of merging bands or transforming the data, the proposed method maintains only those bands that are useful for the hyperspectral image analysis task. The advantage of maintaining physically meaningful bands is to identify the useful wavelengths for a particular classification task. Identifying important wavelengths can also be used in the design of hyperspectral sensors. By reducing the number of wavelengths that need to be collected, data collection will be performed faster and with less required storage space.

Additionally, most of the mentioned algorithms [2]-[5] require the desired number of bands, including Serpico and Bruzzone's search method for band selection, Du, et al.'s method of band prioritization based on the Independent Component Analysis' weight matrix, Wang, et al.'s NFindr volume based method, Han's, et al band prioritization based on eigenvalue weight method, and Guo, et al.'s band selection method based on mutual information [7], [8], [9], [10], and [11]. Often, the number of required bands is unknown.

Keshava presented a method based on the Spectral Angle Mapper (SAM) distance measure [12]. The algorithm incrementally adds bands that increase the SAM measure. Chang, et al. ranks all bands based on loading factors constructed using maximum-variance PCA, MNF, orthogonal subspace projection and minimum misclassification canonical analysis methods [13]. Following ranking, Chang, et al. remove correlated bands using a divergence measure. Chang and Wang's method based on constrained energy minimization selects bands that have the minimal correlation with each other using the concept of virtual dimensionality to determine the number of chosen spectral bands [14]. The proposed method differs from these methods by providing an integrated endmember determination and band selection algorithm. The bands are selected based on their ability to distinguish between endmembers being found rather than using a metric value. The proposed method does weight spectral bands using a method similar to Chang, et al's method of ranking bands according to the Minimum Misclassification Canonical Analysis (MMCA) which is based on Fisher's discriminant function [13]. However, the proposed method differs from Chang, et al's band selection method by performing simultaneous endmember detection and using sparsity promoting priors to prune bands rather than using a divergence measure to determine the number bands.

The proposed band selection method aims to autonomously determine the number of spectral bands required, select bands while performing endmember determination and unmixing, and maintain the physical meaning of the spectral bands. Band selection is integrated into an endmember and abundance determination algorithm by incorporating band weights and a band sparsity term into SPICE's objective function [1].

Section II discusses the sparsity promoting extension of SPICE for band selection. It is assumed that the reader is familiar with the SPICE algorithm, endmember detection and hyperspectral band selection problems. Results are presented in Section III followed by conclusions in Section IV.

## II. Simultaneous Band Selection

In order to perform simultaneous band selection, band weights and a band sparsity promoting term are added to SPICE's objective function [1] which is an extension of the ICE algorithm [15]. Incorporating the band weights and the band sparsity term to the SPICE objective function yields

$$J = \eta \frac{RSS_B}{N} + \beta SSD_B + SPT + BST \tag{1}$$

where

$$RSS_B = \sum_{i=1}^{N} \left( \mathbf{WX}_i - \sum_{k=1}^{M} p_{ik} \mathbf{WE}_k \right)^T \left( \mathbf{WX}_i - \sum_{k=1}^{M} p_{ik} \mathbf{WE}_k \right),$$

$$SSD_B = \sum_{k=1}^{M-1} \sum_{l=k+1}^{M} (\mathbf{WE}_k - \mathbf{WE}_l)^T (\mathbf{WE}_k - \mathbf{WE}_l),$$

$$SPT = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} p_{ik}, \ \mathbf{W} = diag(w_1,...,w_d), \ \gamma_k = \Gamma \Big/ \sum_{i=1}^{N} p_{ik},$$

$w_i$ is the weight for the $i^{th}$ band, $d$ is the number of bands, $\eta$ and $\beta$ are the constant coefficient parameters for the RSS term and SSD terms, $\Gamma$ is a constant controlling the degree of sparsity among the endmembers, and the BST term is the band sparsity promoting term defined below.

The band sparsity promoting term (*BST*) is defined as a weighted sum of band weights, with one term for each band.

$$BST = \sum_{j=1}^{d} \lambda_j |w_j| = \sum_{j=1}^{d} \lambda_j w_j \tag{2}$$

where

$$\lambda_j = \frac{\Lambda \left( \left( \sum_{k=1}^{M} \frac{1}{N} \sum_{i=1}^{N} p_{ik} \left( x_{ij} - e_{kj} \right)^2 \right) + 1 \right)}{\left( \sum_{k=1}^{M} \left( e_{kj} - \mu_{0j} \right)^2 \right) + 1}, \tag{3}$$

$\Lambda$ is a tunable parameter controlling the degree of sparsity among the band weights, $\mu_0$ is the global data mean, $x_{ij}$ is the $j^{\text{th}}$ band of the $i^{\text{th}}$ pixel, and $e_{kj}$ is the $j^{th}$ band of the $k^{th}$ endmember. The band weights are subjected to the following constraints:

$$w_j \geq 0, \ j = 1,....d, \qquad \sum_{j=1}^{d} w_j = d \tag{4}$$

where $d$ is the number of bands. The non-negativity constraint in (4) allows for the second equality in (2).

The $\lambda_j$ values are related to Chang, et al's method of ranking bands according to the Minimum Misclassification Canonical Analysis (MMCA) [13]. In [13], bands are ranked according to the MMCA value which is derived from Fisher's discriminant function. Although the proposed method uses a weight that is related to Chang, et al's effective Fisher's discriminant value, this algorithm differs from their method by performing simultaneous endmember detection and using sparsity promoting priors rather than a divergence threshold to determine the number of useful spectral bands.

Note that, if a $\lambda_j$ value is large, then the associated weight must be small. Hence, a large $\lambda_j$ value should lead to a small band weight for band $j$. The $\lambda_j$ values depend on the ratio of the within-class to between-class scatter. Each endmember has one class consisting of points with high abundances for that endmember. So, bands with small ratios separate the data and endmembers well and are, therefore, encouraged to have large weights. In contrast, bands with large ratios do not separate the data well and are likely pruned.

In order to minimize this new objective function, the iterative procedure in SPICE can be used. The minimization iterates between solving for proportions, endmembers and band weights. The endmembers can be solved for directly as done in [1]. When solving for proportion values given endmember and band weight estimates, N quadratic programming steps, one for each data point, can be employed to minimize (1) with respect to the following constraints.

$$p_{ik} \geq 0, k = 1,....M, \qquad \sum_{k=1}^{M} p_{ik} = 1 \tag{5}$$

Similarly, when solving for band weights given the proportion and endmember estimates, (1) can be minimized using a single quadratic programming step given the constraints in (4). After updating band weights, bands are pruned from the data and the endmembers when the corresponding band weights drop below a prescribed threshold.

Since the band weights and endmember values depend on each other, an optimization schedule needs to be employed. An estimate of the endmembers is needed before determining which bands are useful in distinguishing between the endmembers. Therefore, an update schedule allows the endmembers and proportions to settle before determining band weights. The optimization schedule consists of a starting iteration for band selection, the frequency of band weight updates and a stopping criterion for band weight updates. This iterative schedule is summarized in the pseudo-code below.

```
iter = 1;
BandUpdate = 0;
n = iteration frequency of the band updates
StartBandUpdate = the iteration to start
updating band weights
while (ObjValue - PrevObjValue)^2 > ChgThresh
  Update Proportion Values
  Update Endmember Values
  BandUpdate = 0;
  if (iter > StartBandUpdate)  and (mod_n(iter)
    == 0) and (BandUpdate == 0)
      Update Band Weights
      Prune Bands
      if norm(PrevBandWt-CurrentBandWt)<
      BandChgThresh
        BandUpdate = 1;
      end if
  end if
  Update Objective Function Value, ObjValue
  iter = iter + 1;
end while
```

## III. Results

### A. Simulated Data Set

The B-SPICE algorithm was applied to a simulated data set generated using four normalized endmembers selected from an AVIRIS hyperspectral image of Cuprite, Nevada. The endmembers contained 51 contiguous spectral bands (1978 to 2477 nm) from "Scene 4" of the AVIRIS Cuprite data [16].

The chosen endmembers are shown in Figure 1. The data set is generated from the endmembers using the linear mixing model, that is:

$$\mathbf{X}_i = \sum_{j=1}^{4} p_{ij}\mathbf{e}_j \quad (6)$$

where $\mathbf{X}_i$ is a generated sample point, $\mathbf{e}_j$ is a selected endmember and the $p_{ij}$'s are the generated abundance values.

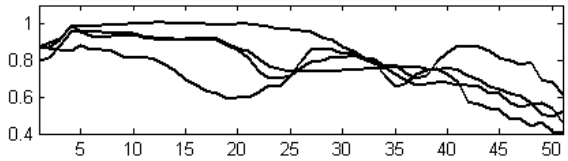

**Figure 1: Endmembers used to generate simulated data set selected by hand from the AVIRIS Cuprite data set.**

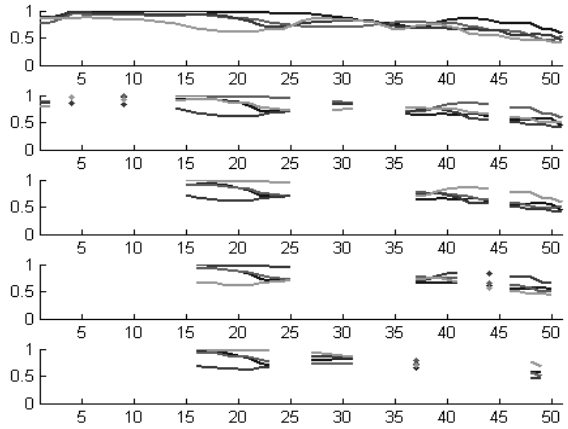

**Figure 2: Endmembers determined using Λ = 0, 0.25, 0.5, 0.75, and 1 on the simulated data set.**

A simulated data set was used to verify that the method can recover the endmembers, perform effective band selection and produce accurate abundance values for each pixel; this can be done since the true endmembers and abundances are known.

B-SPICE and SPICE were run on this data set for a range of Λ values. All parameters, other than those involved with band selection, were held constant for each run of the algorithm. η was set to 2000, β was 0.3, Γ was 0.3 for the first 150 iterations and then set to 0, the initial number of endmembers was set to 20, the endmember pruning threshold was 1e-8, and the initial endmembers were selected randomly from the data set. When running B-SPICE, band selection was not started until the 100th iteration, after which, the band weights were updated every 5th iteration. The band pruning threshold was set to 1e-5 and the band weight change threshold was set to 1e-5. Λ was set to 0 (for SPICE), 0.25, 0.5, 0.75 and 1 (for B-SPICE). B-SPICE and SPICE were run on the data set 50 times for each parameter set. An example of the endmembers found using each Λ value is shown in Figure 2.

Table 1 shows the mean and standard deviation of the number of endmembers and number of bands retained for each parameter set over the 50 runs of the algorithm. As can be seen, both SPICE and B-SPICE are able to consistently determine the correct number of endmembers.

To evaluate the effectiveness of the band selection, the average squared error per abundance value from the true abundance values and the determined abundance values were calculated. These values are shown in Table 2 using the median average squared error per abundance value. The median average squared error per abundance value is computed by taking the median over 50 runs of the algorithm of the average squared error between each pixel's true and computed abundance values. As shown, the median average squared error per abundance value is fairly stable across the Λ values indicating that B-SPICE is as effective at determining the true abundance values as the SPICE algorithm. Therefore, using B-SPICE, the number of bands can be reduced while maintaining the ability to determine abundances. However, when examining the standard deviation of the average squared error per abundance value, it is seen that SPICE is more consistent than B-SPICE. An order of magnitude difference between standard deviations of SPICE and B-SPICE exists.

**Table 1: The mean number and standard deviation of endmembers and bands found over 50 runs of SPICE or B-SPICE on the simulated data set. The true number of endmembers for this data set is 4.**

| Λ | Mean Num. of End-members | Std. Dev. of Num. of End-members | Mean Num. of Bands Retained | Std. Dev. of Num. of Bands Retained |
|---|---|---|---|---|
| 0 (SPICE) | 4 | 0 | 51 | 0 |
| 0.25 | 4 | 0 | 34.6 | 1.1 |
| 0.5 | 4 | 0 | 25.0 | 1.4 |
| 0.75 | 4 | 0 | 20.9 | 1.2 |
| 1 | 4 | 0 | 16.2 | 3.7 |

**Table 2: Statistics of the Average Squared Error per abundance between Calculated and True Abundances**

| Λ | Median Avg. Squared Error per abundance | Mean Avg. Squared Error per abundance | Standard Deviation of Avg. Squared Error per abundance |
|---|---|---|---|
| 0 (SPICE) | 0.005 | 0.005 | 0.0005 |
| 0.25 | 0.004 | 0.008 | 0.0066 |
| 0.5 | 0.004 | 0.007 | 0.0057 |
| 0.75 | 0.004 | 0.007 | 0.0050 |
| 1 | 0.006 | 0.010 | 0.0069 |

### *B. Indian Pines Data Set*

The B-SPICE algorithm was also run on the June 1992 AVIRIS Indian Pines dataset collected in an agricultural area of northern Indiana. The image has 145 x 145 pixels with 220 spectral bands ranging from 0.4 to 2.5 microns. Figure 3 shows band 10 (approximately 0.49 microns) and the ground truth of the dataset. Only 49% of the pixels in the image have ground truth information. The crops in the image are in early growth stages and thus have only about a 5% crop cover. The remaining field area is soil covered with residue from the previous year's crop. The no till, min till, and clean till labels indicate the amount of crop residue remaining from the previous year. No till corresponds to a large amount of residue, min till has a moderate amount and clean till has a minimal amount of residue [7].

SPICE and B-SPICE were run 2 times for 5 different Λ values. All parameters, other than the Λ parameter, were held constant for each run of the algorithm. To reduce run time,

SPICE and B-SPICE were run on 1000 pixels randomly chosen from the dataset. Fast implementations for the algorithm can be created as was done in [15]. After determining the endmembers and selected bands using the subset, unmixing was performed on the entire dataset to find abundance values for every pixel. $\eta$ was set to 5000, $\beta$ to 0.3, $\Gamma$ to 0.2 for the first 100 iterations and then to 0, the initial number of endmembers was set to 20, the endmember pruning threshold was 1e-8, and the initial endmembers were selected randomly from the data set. When running B-SPICE, band selection was not started until the $100^{th}$ iteration, after which, the band weights were updated every $5^{th}$ iteration. The band pruning threshold was set to 1e-5 and the band weight change threshold was set to 1e-5. $\Lambda$ was set to 0 (for SPICE), 0.5, 1, 5, and 10 (for B-SPICE). The number of endmembers and the number of bands found are shown in Table 3.

In order to compare these results to those presented in [11], supervised classification was performed. The features used for supervised classification were the abundance values computed for each pixel in the 16 classes of the dataset. The unlabeled pixels were not included in these experiments. Since the abundance values were the features used for classification, the dimensionality of the feature vectors is equal to the number endmembers found for the dataset.

Two-fold cross validation was performed on the dataset using a 1-vs-1 RVM classification method [17]. The training and testing sets were defined by randomly splitting each of the 16 classes in half. An RVM was trained for each pair of classes. Since there are 16 classes, 120 RVMs were trained for each test set.

Test pixels were classified by counting the number of RVMs that assigned the pixel to each class.

$$L_p = \left[ v_1^p, ..., v_{16}^p \right] \tag{7}$$

where $v_i^p$ is the number of times the pixel $p$ was assigned to class $i$ by the trained RVMs. After every pixel was run through the entire set of trained RVMs, spatial smoothing was performed to assign a label to each pixel. Spatial smoothing was done by summing over the neighborhood of pixel $p$ and assigning the class with the largest number of votes.

$$C_p = \arg\max_{i \in \{1,...,16\}} \{ \sum_{j \in N_p} v_i^j \} \tag{8}$$

where $C_p$ is the label for pixel $p$ and $N_p$ are set of pixels in the 8-connected neighborhood of pixel $p$. The overall classification accuracies for each run of the B-SPICE algorithm are shown in Table 3. Since the classification accuracies depend on the random splitting of the data into training and testing sets, classification was performed three times for each run of the B-SPICE algorithm.

Wang, et al. [9] provide supervised classification results with band selection on the Indian Pines Dataset. Results in [9] show very good classification accuracies ranging from 90% to 94.5% with less than 50 bands; however, their classification method was not described. Band selection results on the Indian Pines dataset can also be found in references [18] and [19] but results are provided on only a subset of the labeled classes. Table 3 also shows results from [20]; only results with less than 50 bands were provided.

### C. Sampling Parameter Values

In order to reduce the need to set parameters by hand, parameters can be sampled from prior distributions. This was implemented by sampling $\eta$, $\beta$, $\Gamma$, and $\Lambda$ from gamma distributions with means of 6000, 0.3, 0.2, and 1. 240 parameters value sets were sampled. B-SPICE was run on the Indian Pines dataset using each of 240 sets of sampled parameters. The results of 240 runs can be combined to determine the number of endmembers, number of bands, and the bands to retain for the dataset. Figure 4 shows the histograms of the number of endmembers, number of bands, and the number of times each band was retained. Modes of the histograms in Figure 4 are 7 and 114, respectively. The most frequently retained bands over the 240 runs were 1-57, 61-76, 81-100, and 118-138. Using these modes and the most frequently retained bands, ICE can be run to find endmembers and abundance values. In other words, the number of endmembers and the bands to retain were determined using the histograms found by running B-SPICE over sampled parameter values. These values were then used to set the number of endmembers and the bands to use for the ICE algorithm. The classification accuracies using sampled parameter values were determined using the same classification method done in the previous Indian Pines experiment in Section III B. Table 4 shows two runs of the ICE algorithm and with three runs of 1-vs.-1 classification.

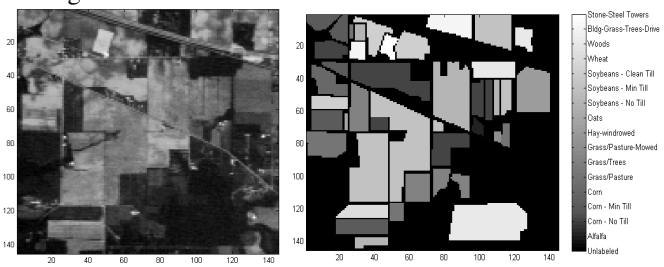


**Figure 3: Band 10 (~0.5 μm) of the AVIRIS Indian Pines Dataset and Ground Truth**

**Table 3: Indian Pines Dataset Results and Comparison. Comparison values estimated from graphs in [11] and [20].**

| Exp. # | Λ | # EMs | # Bands Kept | Classification Accuracy in Percentage | | | Comparison Results in Percentage | |
|---|---|---|---|---|---|---|---|---|
| | | | | Run 1 | Run 2 | Run 3 | Ref. [11] | Ref. [20] |
| 1 | 0 | 8 | 220 | 93.6 | 93.9 | 93.9 | 90 | - |
| 2 | 0 | 7 | 220 | 93.1 | 93.1 | 92.9 | 90 | - |
| 3 | 0.5 | 7 | 124 | 93.3 | 93.7 | 93.7 | 90 | - |
| 4 | 0.5 | 7 | 122 | 93.0 | 92.9 | 93.2 | 90 | - |
| 5 | 1 | 7 | 89 | 93.4 | 93.3 | 93.6 | 90 | - |
| 6 | 1 | 7 | 103 | 93.3 | 93.3 | 93.5 | 90 | - |
| 7 | 5 | 7 | 34 | 86.4 | 86.4 | 86.3 | 88 | 80 |
| 8 | 5 | 8 | 34 | 86.5 | 86.0 | 86.4 | 88 | 80 |
| 9 | 10 | 7 | 19 | 83.4 | 83.9 | 82.5 | 82 | 81 |
| 10 | 10 | 8 | 18 | 77.8 | 80.0 | 78.3 | 82 | 82 |

## IV. Conclusion

The B-SPICE algorithm extends the SPICE algorithm by integrating band selection into the algorithm. The proposed algorithm performs band selection, endmember detection and

unmixing. The proposed algorithm also autonomously determined the number of required endmembers and the number of required spectral bands.

The classification results in Table 3 were generated using the abundance values rather than the reduced band pixels. Because of this, the number of features equals the number of endmembers rather than the number of bands. This is interesting as there are much fewer endmembers than retained bands. Therefore, these classification results have been achieved using a very small number of features.

As noted in Section III, the band selection using sparsity promoting priors is less consistent than endmember detection since standard deviations across the number of bands is higher than across the number of endmembers. Despite the variance in the number of bands, results show good performance.

There are several parameters that need to be set to run the B-SPICE algorithm, e.g., the constant coefficients. Investigation was done into sampling these parameter values rather than setting them directly.

## ACKNOWLEDGMENTS

Research was sponsored by the U. S. Army Research Office and U. S. Army Research Laboratory and was accomplished under Cooperative Agreement Number DAAD19-02-2-0012. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Army Research Laboratory, or the U. S. Government. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

We also thank Leslie Collins and Peter Torrione of Duke University for the use of their DPRT Relevance Vector Machine implementation.

**Table 4: Indian Pines Dataset Results using Sampled Parameter Values and Comparison from [11].**

| Exp. # | # EMs | # Bands Kept | Classification Accuracy in Percentage | | | Comparison Results in Percentage |
|---|---|---|---|---|---|---|
| | | | Run 1 | Run 2 | Run 3 | Reference [11] |
| 1 | 7 | 114 | 92.1 | 92.1 | 92.2 | 90 |
| 2 | 7 | 114 | 92.6 | 92.4 | 92.5 | 90 |

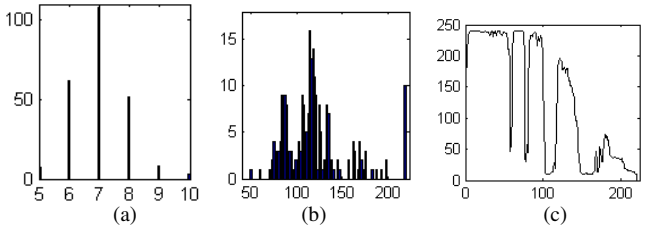


**Figure 4: Histograms of the number of endmembers (a) and bands found (b) and the number of times each band is retained (c) over 240 runs of B-SPICE using sampled parameter values.**

## REFERENCES

[1] A. Zare and P. Gader, "Sparsity Promoting Iterated Constrained Endmember Detection for Hyperspectral Imagery," *IEEE Geosci. and Remote Sens. Lett.*, vol. 4, no. 3, pp. 446-450, Jul. 2007.

[2] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal," *IEEE Trans. Geosci. and Remote Sens.*, vol. 26, no. 1, pp. 65-73, Jan. 1988.

[3] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of High Spectral Resolution Remote-Sensing Data by a Noise-Adjusted Principal Components Transform," *IEEE Trans. Geosci. and Remote Sens.,* vol. 28, no. 3, pp. 295-304, May 1990.

[4] J. C. Harsanyi, and C.-I. Chang, "Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach," *IEEE Trans. Geosci. and Remote Sens.*, vol. 32, no. 4. pp. 779-785, Jul. 1994.

[5] S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, "A Band Selection Technique for Spectral Classification," *IEEE Geosci. and Remote Sens. Lett.*, vol. 2, no. 3, pp. 319-323, Jul. 2005.

[6] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data," *IEEE Trans. Geosci. and Remote Sens.*, vol. 39, no. 7, pp. 1368-1379, Jul. 2001.

[7] S. B. Serpico, L. Bruzzone, "A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 39, no. 7, pp. 1360-1367, Jul. 2001.

[8] H. Du, H. Qi, X. Wang, R. Ramanath, W. E. Snyder, "Band Selection Using Independent Components Analysis for Hyperspectral Image Processing," in *Proc. IEEE 32nd Applied Imagery Pattern Recognition Workshop*, Washington D.C., 2003, pp. 93-98.

[9] L. Wang, Y. Zhang and Y. Gu, "Unsupervised Band Selection Method Based on Improved N-Findr Algorithm for Spectral Unmixing," in *Proc. 1st ISCAA*, Harbin, China, 2006, pp. 1018-1021.

[10] T. Han, D. G. Goodenough, A. Dyk, H. Chen, "Hyperspectral Feature Selection for Forest Classification," in *Proc. IEEE Geosci. and Remote Sens. Symp.*, Anchorage, 2004, pp. 1471-1474.

[11] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band Selection for Hyperspectral Image Classification Using Mutual Information," *IEEE Geosci. and Remote Sens. Lett.*, vol. 3, no. 4, pp. 522-526, Oct. 2006.

[12] N. Keshava, "Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries," *IEEE Trans. Geosci. and Remote Sens.*, vol. 42, no. 7, pp. 1552-1565, Jul. 2004.

[13] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A Joint Band Prioritization and Band Decorrelation Approach to Band Selection for Hyperspectral Image Classification," *IEEE Trans. Geosci. and Remote Sens.,* vol. 37, no. 6, pp. 2631-2641, Nov. 1999.

[14] C.-I. Chang, and S. Wang, "Constrained Band Selection for Hyperspectral Imagery," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 44, No. 6, pp. 1575-1585, June 2006.

[15] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Donne and J. F. Huntington, "ICE: A Statistical Approach to Identifying Endmembers in Hyperspectral Images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 42, pp. 2085–2095, Oct. 2004.

[16] "AVIRIS Free Standard Data Products," [Online]. Sept. 2, 2004, [cited Feb. 2007] Available: http://aviris.jpl.nasa.gov /aviris.freedata.html.

[17] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol. 1, pp. 211-244, Sept. 2001.

[18] R. Archibald and G. Fann, "Feature Selection and Classification of Hyperspectral Images with Support Vector Machines," *IEEE Geosci. and Remote Sens. Lett.*, vol. 4., no. 4, pp. 674-677, Oct. 2007.

[19] R. Huang and M. He, "Band Selection Based on Feature Weighting for Classification of Hyperspectral Data," *IEEE Geosci. and Remote Sens. Lett.*, vol. 2, no. 2, pp. 156-159, Apr. 2005.

[20] A. Martinez-Uso, F. Pla, P. Garcia-Sevilla, and J. M. Sotoca, "Automatic Band Selection in Multispectral Images Using Mutual Information-Based Clustering," in *Proc. 11th Iberoamerican Congress in Pattern Recognition*, Cancun, Nov. 2006, pp. 644-654.