Sparsity Promoting Iterated Constrained Endmember Detection in Hyperspectral Imagery

Alina Zare, Paul Gader *Senior Member, IEEE*

***Abstract*— An extension of the Iterated Constrained Endmembers (ICE) that incorporates sparsity promoting priors to find the correct number of endmembers is presented. In addition to solving for endmembers and endmember fractional maps, this algorithm attempts to autonomously determine the number of endmembers required for a particular scene. The number of endmembers is found by adding a sparsity-promoting term to ICE's objective function.**

***Index Terms*—Sparsity Promotion, Endmember, Hyperspectral Imagery**

I. Introduction

AUTONOMOUS endmember detection is a difficult problem in hyperspectral imaging. Many endmember extraction algorithms have been formulated but the majority of these algorithms require the knowledge of the number of endmembers required for a scene. The problem of autonomously determining the number of required endmembers to a large extent has not been tackled.

We provide an extension of the Iterated Constrained Endmembers (ICE) Algorithm [1] that provides better estimates of the number of endmembers required for a dataset. This extension adds a sparsity-promoting term to the ICE objective function and is, therefore, referred to as SPICE. This added term encourages the pruning of unnecessary endmembers.

In Section II, we review the ICE algorithm and discuss the sparsity promoting extension. It is assumed that the reader is familiar with the endmember detection problem. In Section III, we present results from artificial and real image data. Section IV is the conclusion.

II. ICE with Sparsity Promotion

A. Review of ICE Algorithm

The ICE Algorithm performs a least squares minimization of the residual sum of squares (RSS) based on the convex geometry model. The convex geometry model assumes that every pixel in a scene is a linear combination of the endmembers of the scene. The convex geometry model can be written as

$$\mathbf{X}_i = \sum_{k=1}^{M} p_{ik}\mathbf{E}_k + \varepsilon_i \qquad i = 1,...,N \tag{1}$$

where N is the number of pixels in the image, M is the number of endmembers, ε_i is an error term, p_{ik} is the proportion of endmember k in pixel i, and $\mathbf{E}_k$ is the k^{th} endmember. The proportions satisfy the constraints

$$p_{ik} \geq 0, k = 1,....M, \qquad \sum_{k=1}^{M} p_{ik} = 1. \tag{2}$$

By minimizing the RSS, subjected to the constraints in (2), the error between the pixel spectra and the pixel estimate found by the ICE algorithm for the endmembers and their proportions is minimized.

$$RSS = \sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^T\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right) \tag{3}$$

As described in [1], the minimizer for RSS is not unique. Therefore, the ICE algorithm adds a sum of squared distances (SSD) term to the objective function.

$$SSD = \sum_{k=1}^{M-1}\sum_{l=k+1}^{M}\left(\mathbf{E}_k - \mathbf{E}_l\right)^T\left(\mathbf{E}_k - \mathbf{E}_l\right) \tag{4}$$

This term is proportional to the size of the area bounded by the endmembers. Therefore, by adding this term to the objective function, the algorithm finds endmembers that provide a tight fit around the data. In [1], it is shown that SSD is equivalent to

$$SSD = M(M-1)V \tag{5}$$

where V is the sum of variances (over the bands) of the simplex vertices. As done in [1], V is used in the objective function instead of $M(M-1)V$ in an effort to make this term independent of the number of endmembers, M.

Therefore, the objective function used in the ICE algorithm is

$$RSS_{reg} = (1-\mu)\frac{RSS}{N} + \mu V \tag{6}$$

where μ is the regularization parameter that balances the RSS and SSD terms of the objective function.

The ICE algorithm minimizes this objective function iteratively. First, given endmember estimates, the proportions for each pixel are estimated. For the first iteration of the algorithm, endmember estimates may be set to randomly chosen pixels from the image. This requires a least squares minimization of each term in (3). Since each of these terms is

quadratic and subjected to the linear constraints in (2), the minimization is done using quadratic programming. After solving for the proportions, the endmembers are found using the current proportion estimates:

$$\mathbf{e}_j = \left\{ \mathbf{P}^T\mathbf{P} + \lambda\left(\mathbf{I}_M - \frac{\mathbf{1}\mathbf{1}^T}{M} \right) \right\}^{-1} \mathbf{P}^T x_j \tag{7}$$

where $\mathbf{P}$ is the N x M proportion matrix, $\mathbf{e}_j$ is the vector of endmember values in the j*th* band, $\mathbf{x}_j$ is the vector of all the pixel values in the j*th* band, $\mathbf{I}_M$ is the M x M identity matrix, $\mathbf{1}$ is the M-vector of ones and $\lambda = N\mu/\{(M-1)(1-\mu)\}$. This iterative procedure is continued until the value of RSS_{reg} is smaller than a tolerance value. Although the ICE algorithm is an excellent algorithm for finding endmembers when the number of endmembers is known, there is no automated mechanism in ICE to determine the correct number of endmembers. Our proposed extension uses sparsity promoting priors to alleviate this disadvantage.

B. Sparsity Promotion

The RSS term of the objective function is a least squares term whose minimization is equivalent to the maximization of

$$-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2 = \ln e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2} \quad [2]. \tag{8}$$

When examining the exponential in (8), it can be seen that this is proportional to the Gaussian density with mean $\sum_{k=1}^{M} p_{ik}\mathbf{E}_k$ and variance 1.

$$N\left(\sum_{k=1}^{M} p_{ik}\mathbf{E}_k, 1\right) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2}{2}} \propto e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2} \tag{9}$$

A common method to promote small parameter values during a least squares minimization process is to add a weight decay term to the objective function. The weight decay term attempts to prevent the p_{ik} values from becoming large.

$$LSWD = \ln e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2 - \gamma\sum_{i=1}^{N}\sum_{k=1}^{M} p_{ik}^2}$$
$$= \ln\left[e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2} e^{-\gamma\sum_{i=1}^{N}\sum_{k=1}^{M} p_{ik}^2} \right] \tag{10}$$

where $\gamma \geq 0$. The second exponential in (10) can also be seen as a Gaussian with a mean of zero. Therefore, (10) can be viewed as the log of the following product.

$$p(\mathbf{X}|\mathbf{P})\,p(\mathbf{P}) \tag{11}$$

where $p(\mathbf{X}|\mathbf{P})$ is the probability of the data given the parameters and $p(\mathbf{P})$ is the prior on the parameters.

Unfortunately, the Gaussian prior is not effective at sparsity promotion. The Gaussian does not prefer to set parameter values to zero which would promote sparsity; instead the Gaussian prefers several small valued non-zero parameters. Therefore, instead of using a Gaussian distribution for the parameters' prior, a zero-mean Laplacian distribution can be used to promote sparsity [3].

$$LSSP = -\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x_i} - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^2 - \sum_{k=1}^{M}\gamma_k\sum_{i=1}^{N}|p_{ik}| \tag{12}$$

C. Sparsity Promoting Iterated Constrained Endmembers

Given equation (12), we see that the sparsity promoting term should be of the form:

$$SPT = \sum_{k=1}^{M}\gamma_k\sum_{i=1}^{N}|p_{ik}| = \sum_{k=1}^{M}\gamma_k\sum_{i=1}^{N} p_{ik} \tag{13}$$

where the last equality follows due to the constraints in (2). For this work, we take

$$\gamma_k = \frac{\Gamma}{\sum_{i=1}^{N} p_{ik}}. \tag{14}$$

Γ is a constant associated with the degree that the proportion values are driven to zero. The advantage of this expression for γ_k is that as the proportion values change during the minimization of the objective function, the weight associated with each endmember adjusts accordingly. If the sum of a particular endmember's proportion values becomes small, then the weight, γ_k, for that endmember becomes larger. This weight change accelerates the minimization of those proportion values. Furthermore, since the objective function is minimized in an iterative fashion, the change in the γ_k values does not disrupt the minimization.

Incorporating this new term into ICE's objective function yields

$$RSS^*_{reg} = (1-\mu)\frac{RSS}{N} + \mu V + SPT \tag{15}$$

This can be rewritten as

$$RSS^*_{reg} = \frac{(1-\mu)}{N}\sum_{i=1}^{N}\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^T\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right) + \mu V + \sum_{k=1}^{M}\gamma_k\sum_{i=1}^{N} p_{ik} \tag{16}$$
$$= \frac{(1-\mu)}{N}\sum_{i=1}^{N}\left[\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right)^T\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k \right) + \frac{N}{(1-\mu)}\sum_{k=1}^{M}\gamma_k p_{ik} \right] + \mu V$$

In order to minimize this new objective function, the iterative procedure used in ICE can still be used. The endmembers are still found by solving (7) since SPT does not depend on the endmembers. When solving for the proportion values given endmember estimates, each of the N terms of the

following sum need to be minimized given the constraints in (2) using quadratic programming.

$$RSS^*_{reg\ term1} = \frac{(1-\mu)}{N}\sum_{i=1}^{N}\left[\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^T\left(\mathbf{X}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right) + \sum_{k=1}^{M}\gamma^*_k p_{ik}\right] \quad (17)$$

where

$$\gamma^*_k = \frac{\Gamma^*}{\sum_{i=1}^{N} p_{ik}}, \quad \Gamma^* = \frac{N\Gamma}{(1-\mu)} \quad (18)$$

During the iterative minimization process, endmembers can be pruned as their proportion values drop below a pruning threshold. After every iteration of the minimization process, the maximum proportion values for every endmember can be calculated.

$$MAXP_k = \max_i\{p_{ik}\} \quad (19)$$

If the maximum proportion for an endmember drops below a threshold, then the endmember can be pruned from the endmember set.

III. Results

A. Toy Example

A 2D example was initially used for testing the SPICE algorithm. Figure 1 shows the data set and the endmembers from which the data was generated. The data points were generated in the same fashion as the toy example in [1]. The endmembers used to generate the 100 data points were ($-10\sqrt{2}$, 0), ($10\sqrt{2}$, 0), and (0, 20). The maximum proportions of the bottom two endmembers were 0.80 and 0.60, respectively. Zero-mean, independent Gaussian random noise with a variance of 1 was added to the *x*- and *y*-coordinates of all the data generated.

In [1], the number of endmembers used to solve the simple example was known to be 3. In SPICE, the number of endmembers does not need to be known and, therefore, the initial number of endmembers can be set to a large value.

The results of three experiments comparing the ICE and SPICE algorithms are shown in Figure 2. The parameters for each algorithm, other than the sparsity promoting term, were held the same during the experiments. The initial number of endmembers for all three runs was 20 and μ was set to 0.001. Γ for the SPICE algorithm in these runs were set to 10, 20, and 5, respectively. The endmembers were initialized to the same values for each experiment comparing ICE and SPICE. These initial endmembers were chosen randomly from the data set.

An endmember was pruned from either algorithm when the endmember's maximum proportion over the data points dropped below 0.0005. In these three experiments, the following proportion values were averaged over the iterations

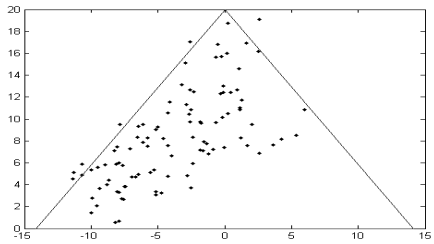

Figure 1 : 2D Example. 100 data points generated from the corners of the simplex shown.

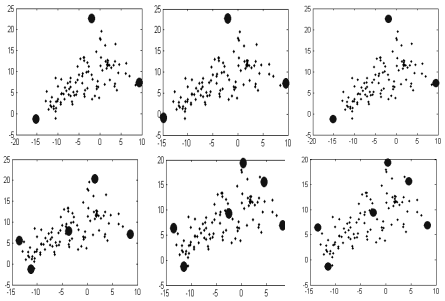

Figure 2: Comparison of SPICE (top) and ICE with pruning (bottom). In these three experiments, μ = 0.001 and the pruning threshold = 0.0005. Initial number of endmembers was 20.

in which an endmember was pruned:

$$MINMAXP_k = \min_k\left\{\max_i\{p_{ik}\}\right\} \quad (20)$$

These were found to be $3.3\text{x}10^{-4}$, $2.4\text{x}10^{-4}$, and $2.4\text{x}10^{-4}$ for ICE, respectively. In comparison, $4.1\text{x}10^{-6}$, $8.3\text{x}10^{-17}$, and $7.8\text{x}10^{-17}$ are these mean values for SPICE in the three experiments, respectively. As can be seen, these values are significantly lower in SPICE compared to the pruning threshold than these values for ICE. SPICE consistently drives proportion values for unnecessary endmembers well below a 0.0005 pruning threshold. Despite this high pruning threshold, ICE did not find the correct number of endmembers with pruning without the use of a sparsity promoting term.

As can be seen in the results shown, SPICE determined that 3 endmembers was an appropriate number to represent the data set. ICE ended the algorithm with 6 endmembers. In the first comparison shown in Figure 2, two of the endmembers found by ICE were (-3.62, 7.94) and (-3.68, 7.94) and so appear to be one endmember in the figure.

B. Cuprite Data Results

To determine SPICE's value on real image data, SPICE was run on AVIRIS hyperspectral image data from Cuprite, Nevada. The data used was 51 contiguous spectral bands (1978 to 2477 nm) from "Scene 4" of the AVIRIS Cuprite data from [4]. This data set was also used in [5]. We chose this data set to be able to compare our results with the NFINDR results presented in [5].

As in [1], SPICE was run on a subset of pixels from the image using "candidate points" selected using the Pixel Purity Index (PPI) [6]. The candidate points in our experiments were

chosen from 10,000 random projections. Points within a distance of 2 from the boundary of the projection received increased purity indices. The 1011 pixels with the highest PPI were used as the candidate points. In [1] , 1000 pixels were used during the experiments on the real image sets. We chose a PPI threshold that allowed us to have as close to 1000 pixels as possible (many pixels have the same PPI).

Results from one experiment on this image are shown in Figure 3. Figure 3 shows the spectral profiles of the nine endmembers found by SPICE. The three endmembers in the first column of Figure 3 compare well to the three endmembers found and identified as Kaolinite, Alunite and Calcite in [5], respectively. Figure 4 shows a comparison of one of the found endmembers to the USGS spectral library data on Alunite [7].

Although it is clear that SPICE was able to find some of the same endmembers identified in [5], it is not clear if the correct number of endmembers was found. The difficulty of using real image data is that the correct number of endmembers in the scene is unknown. To overcome this problem, a subset of the Cuprite data image was used for further testing of the algorithm.

Three endmembers, shown in Figure 5, were selected from the hyperspectral image by hand. The squared Euclidean distance was calculated from every pixel in the image to these three endmembers. The pixels within 500,000 squared Euclidean distance from these three hand-selected endmembers were collected and used as a test set for SPICE. The test set was normalized and is shown in Figure 6.

Table 1 shows the number of endmembers found using SPICE for a range of Γs and initial number of endmembers. As can be seen, SPICE consistently finds three endmembers for this data set. The results in Table 1 and in Figure 2 show that the SPICE algorithm is fairly stable with respect to Γ.

SPICE is also very stable with respect to the initial number of endmembers. Therefore, the initial number of endmembers should be set to a large value.

Figure 7 shows the endmembers that were found by SPICE in these experiments. These endmembers are clearly very similar to the three hand-selected endmembers used for this experiment.

C. Indian Pines Results

SPICE was also run on the June 1992 AVIRIS data set collected over the Indian Pines Test site. The image has 145 x 145 pixels with 220 spectral bands and contains approximately 2/3 agricultural land and 1/3 forest and other elements. The soybean and corn crops in this scene are in early growth stages. Therefore, these regions are primarily bare soil and residues from previous crops [9].

As in the previous experiments, SPICE was run on a subset of the image pixels. 1100 pixels were randomly selected from the image. Before running SPICE, these pixels were normalized. SPICE was initialized to 60 endmembers, μ to 0.1, and Γ to 1. 10 endmembers were found for this dataset using SPICE. The resulting abundance maps are shown in Figure 8. SPICE pruned unnecessary endmembers and provided interpretable results that compare to results found by others on this dataset [8,9,10].

In Figure 8, the images were found to correspond to the following: (a) and (f) correspond to woods and tree canopies,

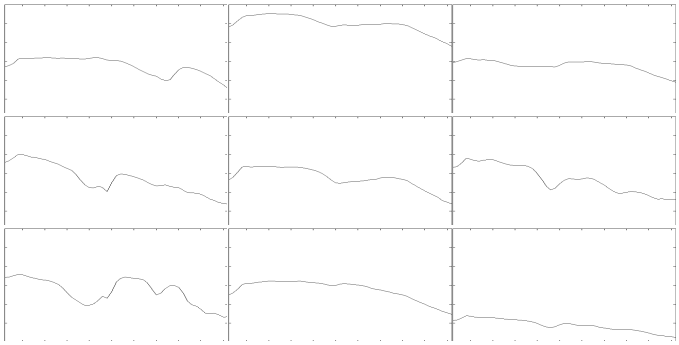

Figure 3: Endmembers found from AVIRIS Cuprite Hyperspectral Data. μ was 0.1 for all experiments. The pruning threshold was set to 1e-9. The limits of the x-axis range are 1978 to 2477 nm and the limits of the y-axis are 1000 to 7000 in units of 10000 times the reflectance factor [7].

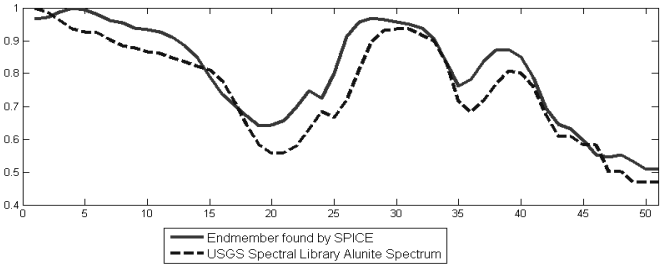

Figure 4: Comparison of one endmember found by SPICE and a USGS Alunite spectrum ("Alunite SUSTDA-20 W1R1Ba AREF") from the 2005 USGS spectral library.

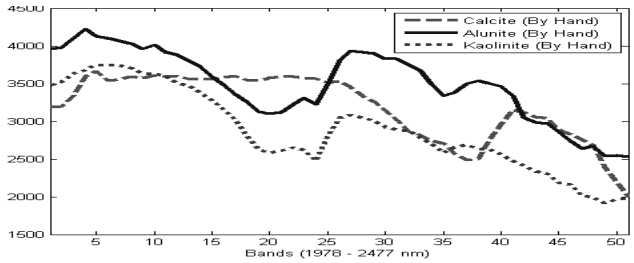

Figure 5: Endmembers selected from AVIRIS Cuprite data image by hand.

(b), (c) and (j) to a mixture of soybean and corn crops, (d) and (e) to grass and background, (f) is hay-windrowed, (g) to steel towers, roads, and other man-made objects, and (h) to grass/pasture and wheat.

IV. Conclusion

The SPICE algorithm extends the ICE algorithm with the addition of a sparsity promoting term. This term encourages the pruning of excess endmembers by penalizing the objective function when a large number of endmembers are being used. The sparsity promoting term drives the set of proportions associated unnecessary endmembers to zero at which point that endmember can be pruned from the set of endmembers representing the data.

Although results suggest that the SPICE algorithm removes

the need to know the number of endmembers needed for a scene in advance, there are still a number of parameters that need to be set, e.g., the gamma constant and the regularization parameters. Future work can include investigation of methods to automatically set these parameters.

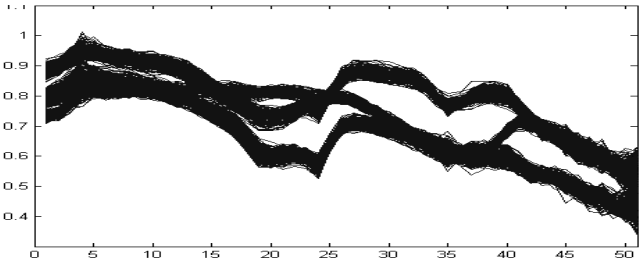

Figure 6: Normalized Test Pixels selected from Cuprite data.

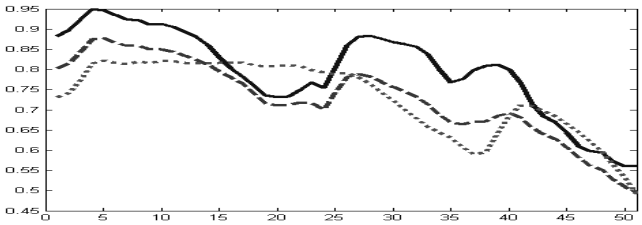

Figure 7: SPICE Endmember Results found on normalized test data selected from the Cuprite AVIRIS scene.

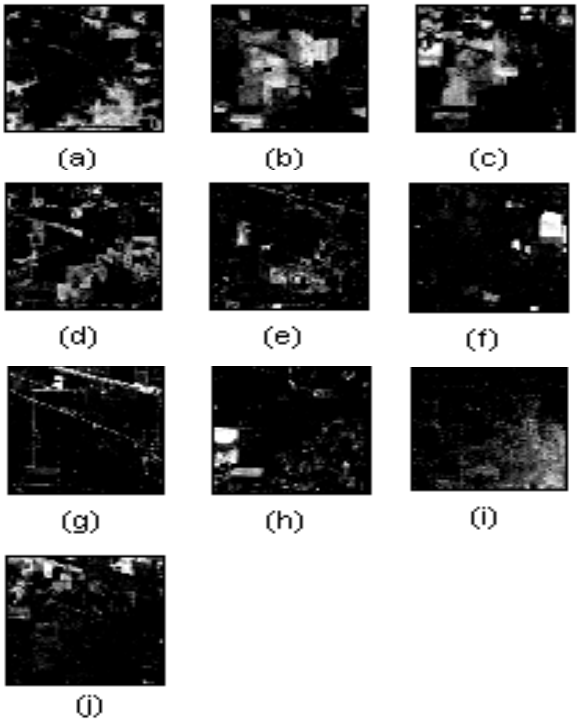

Figure 8: Abundance maps generated by SPICE on the Indian Pines data set.

ACKNOWLEDGMENT

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.

Exp.	Initial # of EM	Gamma const. for SPICE	# of found EM, SPICE	# of found EM, ICE
1	5	1	3	5
2	10	0.5	3	9
3	10	0.5	3	8
4	10	10	3	9
5	10	10	3	8
6	15	1	3	12
8	30	1	3	12
9	40	1	3	13
10	50	1	3	11

Table 1: The number of endmembers found by SPICE and ICE on Test Pixels from AVIRIS Cuprite Data over a range of Γ values and initial number of endmembers. Each experiment had the same initialization for ICE and SPICE. The μ parameter was 0.1 for all experiments. The pruning threshold was set to 1e-9.

REFERENCES

[1] 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. On Geoscience and Remote Sensing*, vol. 42, Oct. 2004, pp. 2085–2095.

[2] P. Williams, "Bayesian Regularization and Pruning Using a Laplace Prior," *Neural Computation*, vol. 7, pp. 117-143, 1995.

[3] M. A. T. Figueiredo, "Adaptive Sparseness for Supervised Learning," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 25, Sept. 2003, pp. 1150–1159.

[4] Jet Propulsion Laboratory AVIRIS website: http://aviris.jpl.nasa.gov/.

[5] Winter, M. E., "Fast Autonomous Spectral End-member Determination In Hyperspectral Data", *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, Vol. II, Vancouver, B.C., Canada, 1999, pp 337-344.

[6] J. Boardmann, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," *Summaries of the 5th Annu. JPL Airborne Geoscience Workshop*, vol. 1, AVIRIS Workshop, R. Green, Ed., Pasadena, CA, 1995, JPL Publ. 95-1, pp 23-26.

[7] USGS Spectroscopy Lab website: http://speclab.cr.usgs.gov/.

[8] L. Miao, H. Qi, and H. Szu, "Unsupervised Decomposition of Mixed Pixels Using the Maximum Entropy Principle," *Proceedings of the 18th International Conference On Pattern Recognition*, Vol. 1, 2006, pp 1067-1070.

[9] M. Grana, M.J. Gallego, "Associative Morphological Memories for Spectral Unmixing," *European Symposium on Artificial Neural Networks*, April 2003, pp 481-486.

[10] M. Grana, P. Sussner, G. Ritter, "Associative Morphological Memories for Endmember Detection in Spectral Unmixing," *The 12th IEEE International Conference on Fuzzy Systems*, Vol. 2, May 2003, pp. 1285-1290.