Sparsity Promoting Iterated Constrained Endmember Detection in Hyperspectral Imagery

Alina Zare and Paul Gader, *Senior Member, IEEE*

***Abstract*—An extension of the iterated constrained endmember (ICE) algorithm 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 that are required for a particular scene. The number of endmembers is found by adding a sparsity-promoting term to ICE's objective function.**

***Index Terms*—Endmember, hyperspectral imagery, sparsity promotion.**

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 knowledge of the number of endmembers that are 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 endmember (ICE) algorithm [1] that provides better estimates of the number of endmembers that are required for a data set. This extension adds a sparsity-promoting term to the ICE objective function and is therefore referred to as sparsity-promoting ICE (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 presents the conclusion.

II. ICE With Sparsity Promotion

A. Review of the 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

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

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 kth endmember. The proportions satisfy the constraints

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

By minimizing the RSS, which is 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, i.e.,

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

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, i.e.,

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

This term is proportional to the size of the area that is 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

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

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

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

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

Manuscript received October 25, 2006; revised February 5, 2007. This work was supported by the U.S. Army Research Office and the U.S. Army Research Laboratory, and was accomplished under Cooperative Agreement DAAD19-02-2-0012.

The authors are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611-6120 USA.

Digital Object Identifier 10.1109/LGRS.2007.895727

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, i.e.,

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

where $\mathbf{P}$ is the $N \times M$ proportion matrix, $\mathbf{e}_j$ is the vector of the endmember values in the jth band, $\mathbf{x}_j$ is the vector of all the pixel values in the jth band, $\mathbf{I}_M$ is the $M \times M$ 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 the following [2]:

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

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

$$N\left(\sum_{k=1}^{M} p_{ik}\mathbf{E}_k, 1\right) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\sum_{i=1}^{N}\left(\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2}{2} \right] \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{E}_k\right)^2 \right]. \tag{9}$$

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, i.e.,

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

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 nonzero parameters. Therefore, instead of using a Gaussian distribution for the parameters' prior, a zero-mean Laplacian distribution can be used to promote sparsity [3], i.e.,

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

C. SPICE

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

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

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

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

Γ is a constant that is 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 that is associated with each endmember adjusts accordingly. If the sum of a particular endmember's proportion values becomes small, then 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

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

This can be rewritten as

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

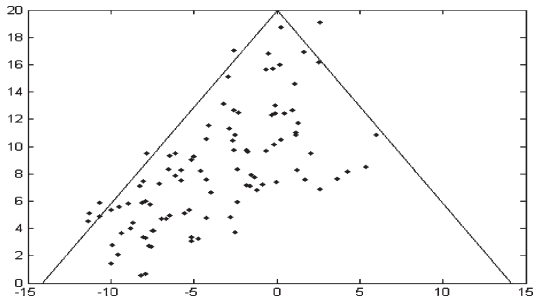

Fig. 1. Two-dimensional example. One hundred data points that are generated from the corners of the simplex are shown.

In order to minimize this new objective function, the iterative procedure that is 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:

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

where

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

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 as

$$\mathrm{MAX}P_k = \max_i\{p_{ik}\}. \tag{19}$$

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

III. Result

A. Toy Example

A 2-D example was initially used for testing the SPICE algorithm. Fig. 1 shows the data set and the endmembers from which the data were generated. The data points were generated in the same fashion as the toy example in [1]. The endmembers that are 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 generated data.

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

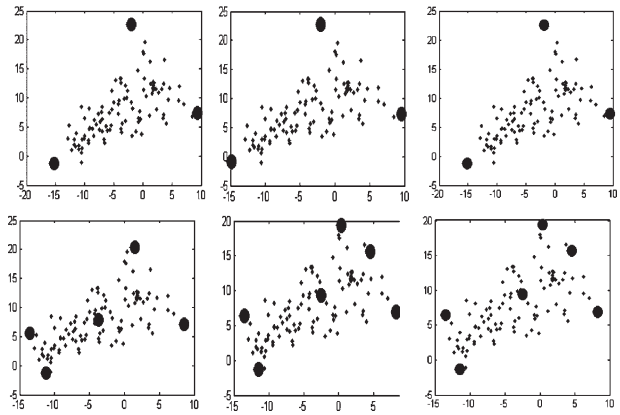

Fig. 2. Comparison of (top) SPICE and (bottom) ICE with pruning. In these three experiments, $\mu = 0.001$, and the pruning threshold is 0.0005. The initial number of endmembers was 20.

The results of three experiments comparing the ICE and SPICE algorithms are shown in Fig. 2. The parameters for each algorithm, other than the sparsity-promoting term, were set to be 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 was 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 in which an endmember was pruned:

$$\mathrm{MINMAX}P_k = \min_k\left\{\max_i\{p_{ik}\}\right\}. \tag{20}$$

These were found to be 3.3×10^{-4}, 2.4×10^{-4}, and 2.4×10^{-4} for ICE, respectively. In comparison, 4.1×10^{-6}, 8.3×10^{-17}, and 7.8×10^{-17} are these mean values for SPICE in the three experiments, respectively. As shown, these values are significantly lower in SPICE compared to the pruning threshold than the values in 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 shown by the results, SPICE determined that three endmembers was an appropriate number to represent the data set. ICE ended the algorithm with six endmembers. In the first comparison that is shown in Fig. 2, two of the endmembers that are found by ICE were $(-3.62, 7.94)$ and $(-3.68, 7.94)$, and they appear as one endmember in the figure.

B. Cuprite Data Results

To determine SPICE's value on real image data, SPICE was run on Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral image data from Cuprite, NV. The

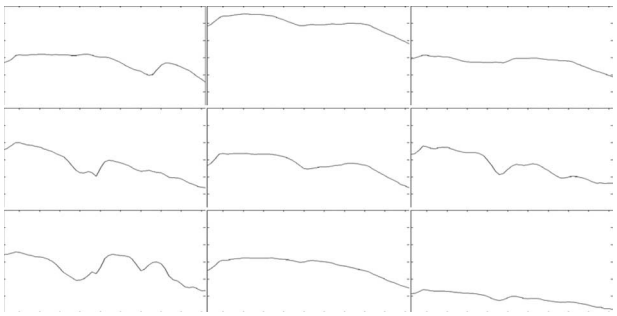

Fig. 3. Endmembers that are found from the AVIRIS Cuprite hyperspectral data. μ was 0.1 for all experiments. The pruning threshold was set to $1e^{-9}$. The limits of the x axis are in the range of 1978–2477 nm, and the limits of the y axis are in the range of 1000–7000 in units of 10 000 times the reflectance factor [7].

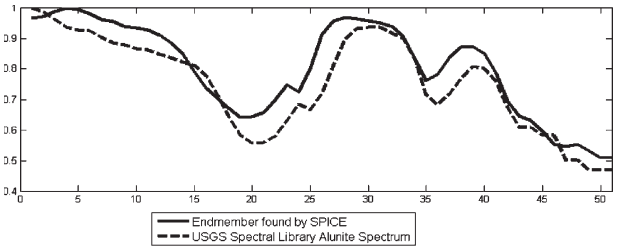

Fig. 4. Comparison of one endmember that is found by SPICE and a USGS Alunite spectrum ("Alunite SUSTDA-20 W1R1Ba AREF") from the 2005 USGS spectral library.

data used were 51 contiguous spectral bands (in the range of 1978–2477 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 that are presented in [5].

As in [1], SPICE was run on a subset of pixels from the image using "candidate points" that are 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 two 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 and the spectral profiles of the nine endmembers that are found by SPICE are shown in Fig. 3. The three endmembers in the first column of Fig. 3 compare well to the three endmembers that are found and identified as kaolinite, alunite, and calcite in [5], respectively. Fig. 4 shows a comparison of one of the found endmembers to the U.S. Geological Survey (USGS) spectral library data on alunite [7].

Although it is clear that SPICE was able to find some of the same endmembers that are 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.

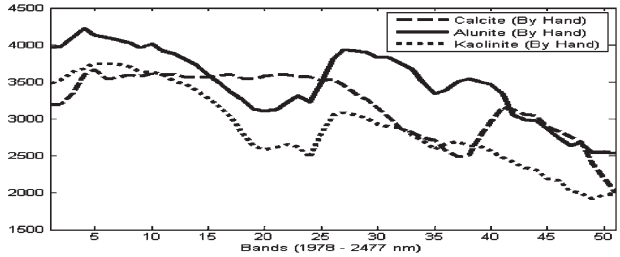

Fig. 5. Endmembers that are selected from AVIRIS Cuprite data image by hand.

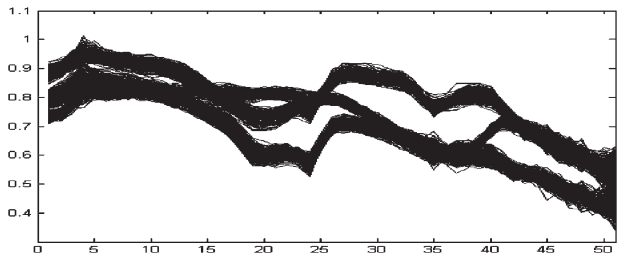

Fig. 6. Normalized test pixels that are selected from the Cuprite data.

TABLE I
NUMBER OF ENDMEMBERS THAT ARE FOUND BY SPICE AND ICE ON TEST PIXELS FROM AVIRIS CUPRITE DATA OVER A RANGE OF Γ VALUES, AND THE INITIAL NUMBER OF ENDMEMBERS. EACH EXPERIMENT HAD THE SAME INITIALIZATION FOR ICE AND SPICE. μ WAS SET TO 0.1 FOR ALL EXPERIMENTS. THE PRUNING THRESHOLD WAS SET TO $1e^{-9}$

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

Three endmembers, as shown in Fig. 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 Fig. 6.

Table I shows the number of endmembers that are found using SPICE for a range of Γs and an initial number of endmembers. As shown, SPICE consistently finds three endmembers for this data set. The results in Table I and in Fig. 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.

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

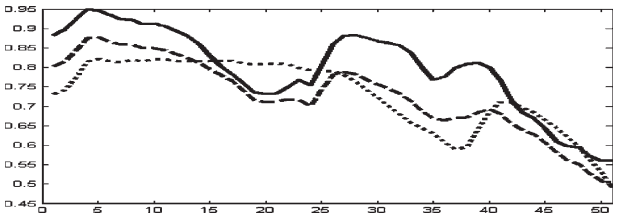

Fig. 7. SPICE endmember results that are found on normalized test data that are selected from the Cuprite AVIRIS scene.

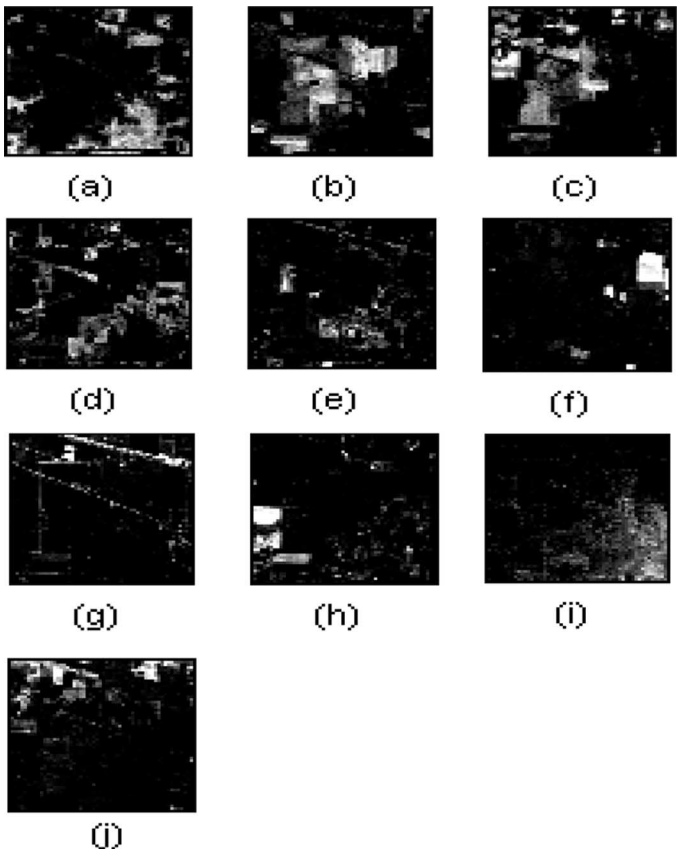

Fig. 8. Abundance maps that are generated by SPICE on the Indian Pines data set.

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 145×145 pixels with 220 spectral bands and contains approximately two-third agricultural land and one-third forest and other elements. The soybean and corn crops in this scene are in their 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. One thousand one hundred pixels were randomly selected from the image. Before running SPICE, these pixels were normalized. SPICE, μ, and Γ were initialized to 60 endmembers, 0.1, and 1, respectively. Ten endmembers were found for this data set using SPICE. The resulting abundance maps are shown in Fig. 8. SPICE pruned unnecessary endmembers and provided interpretable results that compare to results that are found by others on this data set [8]–[10].

In Fig. 8, the images were found to correspond to the following: (a) and (f) woods and tree canopies; (b), (c), and (j) a mixture of soybean and corn crops; (d) and (e) grass and background; (f) windrows; (g) steel towers, roads, and other man-made objects; and (h) 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 that are associated with unnecessary endmembers to zero, at which point such endmembers 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 that are 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.

Acknowledgment

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.

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. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[2] P. Williams, "Bayesian regularization and pruning using a Laplace prior," *Neural Comput.*, vol. 7, no. 1, pp. 117–143, Jan. 1995.

[3] M. A. T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1150–1159, Sep. 2003.

[4] *AVIRIS Free Standard Data Products*. (2004, Sep. 2). [Online]. Available: http://aviris.jpl.nasa.gov/aviris.freedata.html

[5] M. E. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, Vancouver, BC, Canada, 1999, vol. 2, pp. 337–344.

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

[7] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley. (2004, Jan. 5). *USGS Digital Spectral Library (splib05a)*. [Online]. Available: http://speclab.cr.usgs.gov/spectral-lib.html

[8] L. Miao, H. Qi, and H. Szu, "Unsupervised decomposition of mixed pixels using the maximum entropy principle," in *Proc. 18th Int. Conf. Pattern Recog.*, 2006, vol. 1, pp. 1067–1070.

[9] M. Grana and M. J. Gallego, "Associative morphological memories for spectral unmixing," in *Proc. Eur. Symp. Artif. Neural Netw.*, Apr. 2003, pp. 481–486.

[10] M. Grana, P. Sussner, and G. Ritter, "Associative morphological memories for endmember detection in spectral unmixing," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst.*, May 2003, vol. 2, pp. 1285–1290.