Comparison of Approaches for Determining End-members in Hyperspectral Data [1,2]

Michael E. Winter University of Queensland, St. Lucia, Queensland, Australia
winter@shake2.earthsciences.uq.edu
Edwin M. Winter Technical Research Associates, Inc., 760 Las Posas Rd. Suite A-4 Camarillo, CA 93010
edwinter@tracam.com

Abstract—A very useful analysis approach for hyperspectral data has been linear unmixing which is a projection into a coordinate system where the coordinates are the constituent or endmember spectra of the scene. The most useful technique is to determine the spectra from the image. Once these spectra are found, the image cube can be "unmixed" into fractional abundances of each material in each pixel. Several autonomous methods, ORASIS, the first real time autonomous algorithm; N-FINDR, an algorithm that determines endmembers by inflating a simplex, and Iterative Error Analysis (IEA) which finds endmembers by iterative unmixing are compared in this paper. Computer implementations of the three autonomous algorithms are evaluated using hyperspectral AVIRIS data of Cuprite, Nevada. The fractional abundance maps produced by the three algorithms are compared to a mineral map made by the USGS based upon an AVIRIS scene, and the end-member spectra are compared to library spectra.

TABLE OF CONTENTS

1. INTRODUCTION

New airborne and satellite hyperspectral sensors are being developed for remote sensing applications. In the near future, these sensors will be producing a near continual stream of high dimensional data. This expected high data volume will require faster means for analysis. A very useful analysis approach has been to project the hyperspectral data into a coordinate system where the coordinates are the constituent spectra of the scene. This procedure is linear unmixing and involves solving the inverse problem to project the data cube onto the spectra. While the spectra can be from a library, the most useful technique is to determine them from the image. These spectra are called "end-members" and a variety of methods have been proposed to find them in the data cube [1]. The Pixel Purity Index (PPI) algorithm [1] first finds the "purest" pixels in the scene through a procedure of repeated projections on to randomly oriented lines in N-dimensional space. These pure pixels are then reviewed by a trained analyst using a method of N-dimensional visualization to select scene end-member spectra. This procedure, based on the geometry of convex sets, has been successfully used for some time, but it is highly interactive. There exists a need for procedures that can be used to process large data sets rapidly, to select specific data sets for interactive review.

Several new autonomous methods for finding these end-members in the hyperspectral data cube and then unmixing the data will be discussed in this paper. These approaches all use different methods to find the end-members in the hyperspectral data cubes. The Orasis algorithm, the first proposed real-time autonomous method [3], uses a "shrink-wrap" approach to find the end-members. The N-FINDR algorithm [4] finds the simplex of maximum volume that can be enclosed within all of the data points. The Iterative Error Analysis (IEA) approach [5] finds the end-members by doing a series of constrained unmixings and choosing as the end-members those pixels that minimize the error in the unmixed image. In addition to these physically based methods, there are several statistically based approaches, such as the Stochastic Expectation Maximization (SEM) approach. SEM is a spectral clustering technique for classifying spectral terrain data that involves iterative estimation of a Gaussian mixture model fit to spectral data [6].

---

[1] 0-7803-5846-5/00/$10.00 © 2000 IEEE

[2] Research performed at Technical Research Associates, Inc.

This paper will concentrate on the physically based methods that utilize the spectra themselves. The procedure for unmixing the hyperspectral data cube is similar for all of these approaches. The points are unmixed using a least squares technique into the constituent fractional abundance planes, which tell the fraction of each pixel that is composed of each constituent. Determining the end-members from the image and then unmixing them reduces a hyperspectral image of large numbers of bands to a much more concise image with a few physically meaningful bands. This determination of the end-members and the subsequent unmixing can be accomplished without any *a priori* knowledge of the image.

2. THEORETICAL BACKGROUND

Generally, the spectra for a given pixel in an image is assumed to be a linear combination of the end-member spectra:

$$p_{ij} = \sum_k e_{ik} c_{kj} + \varepsilon \quad , \qquad (1)$$

where p_{ij} is the i-th band of the j-th pixel, e_{ik} is the i-th band of the k-th end-member, c_{kj} is the mixing proportions for the j-th pixel from the k-th end-member, and ε is gaussian random error (assumed to be small). Since the pixel compositions are assumed to be percentages, the mixing proportions are assumed to sum to one:

$$\sum_k c_{kj} = 1 \quad , \qquad (2)$$

If the above mixing model is correct, and the number of image bands is exactly one less than the number of end-members, the pixels in an image occupy a space formed by a simplex (see Figure 1). A simplex is the simplest geometric shape that can enclose a space of a given dimension. Examples include the familiar line (one dimension), triangle (two dimensions) and tetrahedron (three dimensions). The vertices of this simplex are the end-member spectra (one of which is a "shade" point which specifies the sensor's response to a completely dark pixel). Given the number of pixels in a large image, it is likely that some pixels will be unmixed, or pure. These pixels are composed of only one of the end-members. The set of points specifying the vertices of the simplex (the end-member set) has the following properties: 1). They represent the set of points in the data that contain the largest "volume"; 2). The same pixels are end-members if the entire data cloud is translated or scaled; 3). The identity of the end-members does not change under any rotation. The four algorithms (Pixel Purity Index, Orasis, N-FINDR and IEA) all search for pure pixels in an image cube, which is the same as finding the vertices of the simplex containing the data. The assumption that a hyperspectral image can be represented as a simplex forms the basis of the geometrical interpretations of hyperspectral data sets [7].

The fact that rotation of the data cloud does not change the identity of the end-members is of particular importance for several of the algorithms, since it allows the use of orthogonal subspace projections (OSPs) for the reduction of the image dimensions. Dimension reduction via OSP (i.e. Principal Components (PC) transform, Minimum Noise Fraction (MNF) transform, or singular value decomposition) is used to reduce the number of bands in an image and simplify the subsequent processing. In this dimensionality reduction step, the number of bands actually processed by the algorithm can be reduced from a hundred or more to the first 15 to 30 OSP coordinates. By discarding the remaining OSP components, there is little information lost, except if small objects are of interest. In the case of spatially small objects, consisting of a few pixels, meaningful information is often contained in the higher components.

3. AUTONOMOUS ALGORITHMS

In the following section, the procedures of the different algorithms are discussed further. There are major differences between the algorithms in the way they find the end-members. In addition, several algorithms choose to reduce the volume of the data either through a dimensionality reduction through sub-space projection or by thinning the data cloud volume through the calculation of exemplar vectors.

Pixel Purity Index (PPI)

A dimensionality reduction is first performed using the MNF transform. The next step is the calculation of the pixel purity index for each point in the image cube. This is accomplished by randomly generating lines in the N-dimensional space comprising a scatter plot of the MNF transformed data. All of the points in the space are then projected onto the line. Those pixels that fall at the extremes of the lines are counted. After many repeated projections to different lines, those pixels with a count above a certain threshold are declared "pure". As a result, there will be many redundant spectra in the pure pixel list. The actual end-member spectra are selected by a combination of

intelligent review of the spectra themselves and through N-dimensional visualization. The fractional abundance maps are found by a non-negative least squares solution of equations 1 and 2.

Orasis

Even though Orasis is tuned for rapid execution, there is no dimensionality reduction as part of this algorithm. Orasis uses a process called Exemplar Selection to thin the data set [8]. The procedure rejects any "redundant" spectra, by calculating the angle between spectral vectors. Any vector not separated by a threshold angle is thinned from the data. This threshold angle is related to the sensor noise and if chosen carefully will not result in the loss of any important end-member. The Orasis algorithm then finds a basis set of much lower dimension, than the original data, by a modified Gram-Schmidt process. The exemplar spectra are then projected onto this basis subspace and a simplex found through a minimum volume transform [1]. To maintain its ability to process sensor data in real-time, speed, Orasis uses filter vectors, an alternative method to a least squares solution without the constraint of requiring the abundances to lie between zero and one, to calculate the fractional abundance maps.

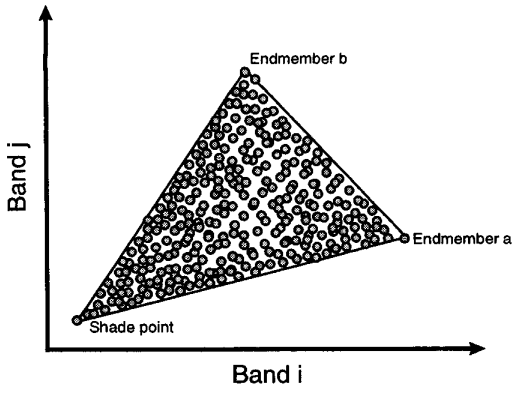


Figure 1. Scatter plots between two bands typically show a triangular shape, with the data radiating away from the shade point.

N-FINDR

This algorithm was originally formulated to work on data after dimensionality reduction via sub-space projection. Dimensionality reduction using sub-space projection (followed by the discard of the higher components) favors abundant scene components. Because this means that localized spectral anomalies may be missed, a later implementation has bypassed this step. The input to this process is the full spectral image cube, without either the dimensionality reduction or the exemplar thinning process. The procedure must examine the full data set to find those pure pixels that can be used to describe the various mixed pixels in the scene. This algorithm finds the set of pixels with the largest possible volume by "inflating" a simplex inside the data. The procedure begins with a random set of vectors. In order to refine the estimate of the end-members, every pixel in the image must be evaluated as to its likelihood of being a pure or nearly pure pixel. To do this, the volume must be calculated with each pixel in place of each end-member. A trial volume is calculated for every pixel in each end-member position by replacing that end-member and finding the volume. If the replacement results in an increase in volume, the pixel replaces the end-member. This procedure is repeated until there are no more replacements of end-members. Once the end-members are found, their spectra can be used to unmix the original image using either linear inversion or non-negatively constrained least squares. This produces a set of images, each of which shows the fractional abundance of an end-member in each pixel. The fractional abundance maps are found by a non-negative least squares solution of equations 1 and 2. Computer time is also a consideration. While the end-member determination step of N-FINDR has been optimized and can be executed rapidly, the unmix step can be time consuming. The constrained unmix can take up to four times longer than the unconstrained least squares unmix (similar to the filter vector approach of Orasis).

Iterative Error Analysis (IEA)

This algorithm is also executed directly on the data with neither the dimensionality reduction nor the data thinning. An initial vector (usually the mean spectrum of the data) is chosen to start the process. A constrained unmix is then performed and the error image is then formed. The average of the vectors with the largest error (distance from the initial vector) is assumed to be the first end-member. Another constrained unmix is then performed and the error image formed. The average of the vectors with the largest error (distance from the first end-member) is assumed to be the second end-member. This process is continued until the predetermined number of end-members is found. The fractional abundance maps are found by the final constrained unmix [9]. There are no published computer time estimates for the IEA algorithm, but the repeated constrained unmixes should take significant time.

4. APPLICATION OF N-FINDR

In this section, the N-FINDR algorithm is applied to hyperspectral data collected by the AVIRIS sensor over Cuprite, Nevada is discussed. AVIRIS is a high quality low noise hyperspectral instrument that acquires data in 224 contiguous spectral bands from 400 to 2500 nm at approximately 20-meter resolution over a swath width of 10-km [10]. Cuprite is a mining area in southern Nevada with hydrothermally altered and unaltered rocks and little vegetation. It has been extensively used for remote sensing experiments over the past twenty years and the minerals have been extensively mapped. For this N-FINDR experiment, 50 contiguous SWIR bands (1978 to 2478 nm) were selected from the 1996 AVIRIS flight data. A ten by ten-kilometer area covering the geologically interesting parts of the AVIRIS scene was processed by N-FINDR. A broadband SWIR image of the selected scene is shown in Figure 2.

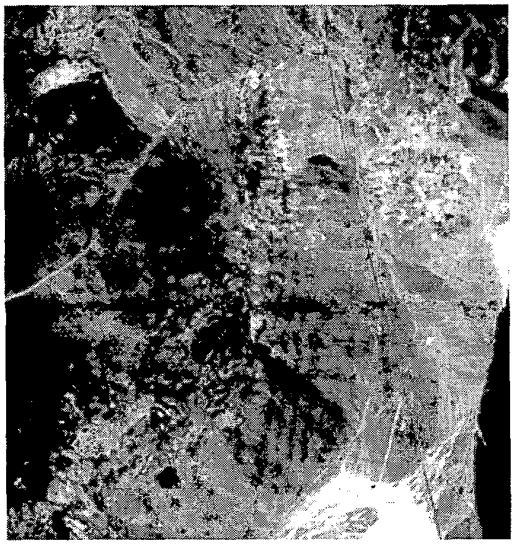


Figure 2. Broad band SWIR Image Formed by Summing the Fifty Spectral Bands Used for Processing

Through efficient implementation of the algorithm, the full process of autonomous end-member determination and spectral unmixing was accomplished in approximately 30 seconds using a Pentium II 400 MHz computer with 512 Megabytes of RAM. Some of the results of the processing are shown in Figures 3 and 4. Four of the fifteen derived fractional abundance maps are shown in Figure 3. The fractional abundance maps depict the fractional composition of the end-member material as a gray level image. Thresholded mineral maps for the same three minerals based upon the 1995 AVIRIS scene are also shown. Since the USGS mineral maps are thresholded and do not include mixtures, there is not a one-to-one correspondence. However, the locations of the minerals found by the N-FINDR algorithm are similar to the USGS map. The corresponding end-member spectra for the three fractional abundance images derived using the N-FINDR algorithm are shown in Figure 4. These spectra were then compared to the USGS library [11] in Figure 5 and identified as alunite, kaolinite and calcite.

5. COMPARISON OF ALGORITHMS

There have been some comparisons of the algorithms to laboratory spectra and alternative unmixing methods done by the algorithm authors. For example, the output of Orasis is compared to the Pixel Purity Method using the Cuprite AVIRIS scene in Reference [8], the result of applying the IEA method to high resolution Cuprite data is discussed in Reference [9], and the application of N-FINDR to the Cuprite scene is discussed above. In general, each algorithm finds appropriate spectra for end-members. In this next section, the output of Orasis and our implementation of the IEA method based upon the description of the algorithm [9] will be compared to the output of N-FINDR on the same Cuprite data set. Our implementation of the IEA algorithm will be denoted Statistical Expansion (SE) to avoid confusion with the actual IEA algorithm. Qualitatively it appears to produce results similar to those discussed in Reference [9], but we do not claim it to be the same. Although the three algorithms find end-members using different techniques, the output of the algorithms operating on the same data set is very similar. The end-member spectra for N-FINDR, Orasis and SE derived spectra are shown in Figure 5 for alunite, kaolinite and calcite. Note that there are some subtle differences. The N-FINDR and SE spectra have the same units as the input data, since each end-member is actually the spectrum of a real image pixel. The Orasis end-member spectra have an arbitrary scaling. It must be pointed out that each algorithm (and any other automated algorithm) also produces false end-member spectra due to noise and sensor artifacts. The fractional abundance planes for the three minerals are shown in Figure 6. The N-FINDR and SE abundance images are scaled between zero and one, while the Orasis fractional abundance images have arbitrary units. The differences between the abundance images can be attributed to the fact that the filter vector method used by Orasis to unmix is not constrained.

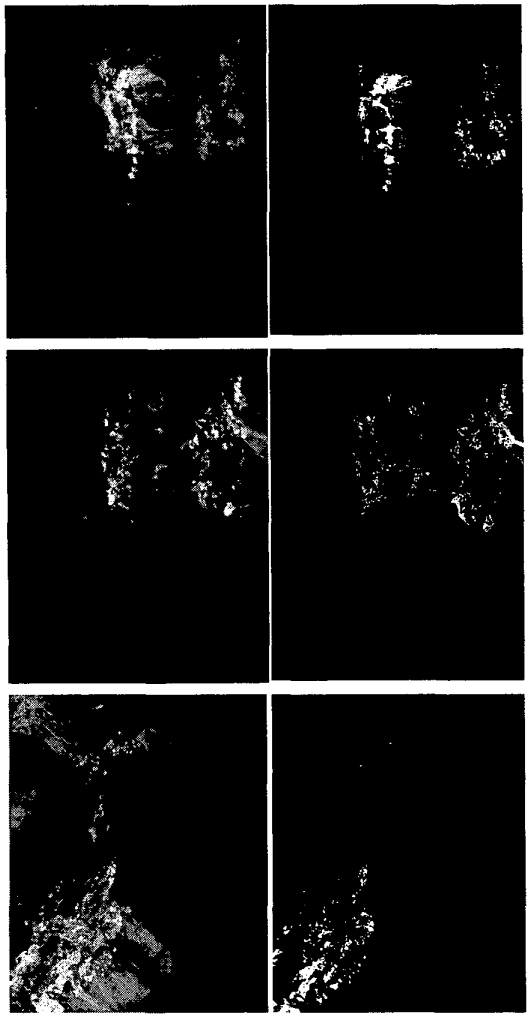


Figure 3. N-FINDR Algorithm Derived Gray Scale Abundance Maps for Alunite, Kaolinite, Calcite and Buddingtonite Compared to USGS Mapping of Minerals [12]. In these gray level abundance maps, white corresponds to individual pixel abundances of 50% or more of the end-member constituent. The individual mineral maps on the right from the USGS composite mineral map were formed by thresholding individual colors

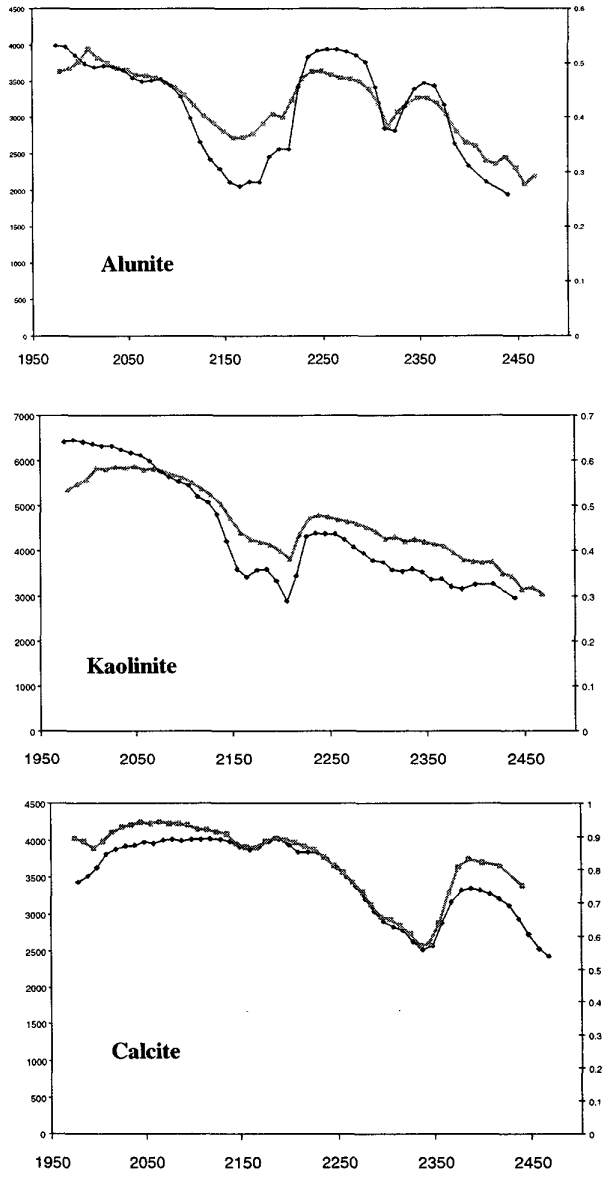


Figure 4. Algorithm Scene Derived (thin red lines) and USGS Laboratory Reflectance (thick blue lines) Spectra for Alunite, Kaolinite and Calcite. Units are reflectance times 10000 for the AVIRIS derived reflectances. Note that there are some differences in the absolute value of the reflectances and spectral shape that are likely due to atmospheric transmission effects.

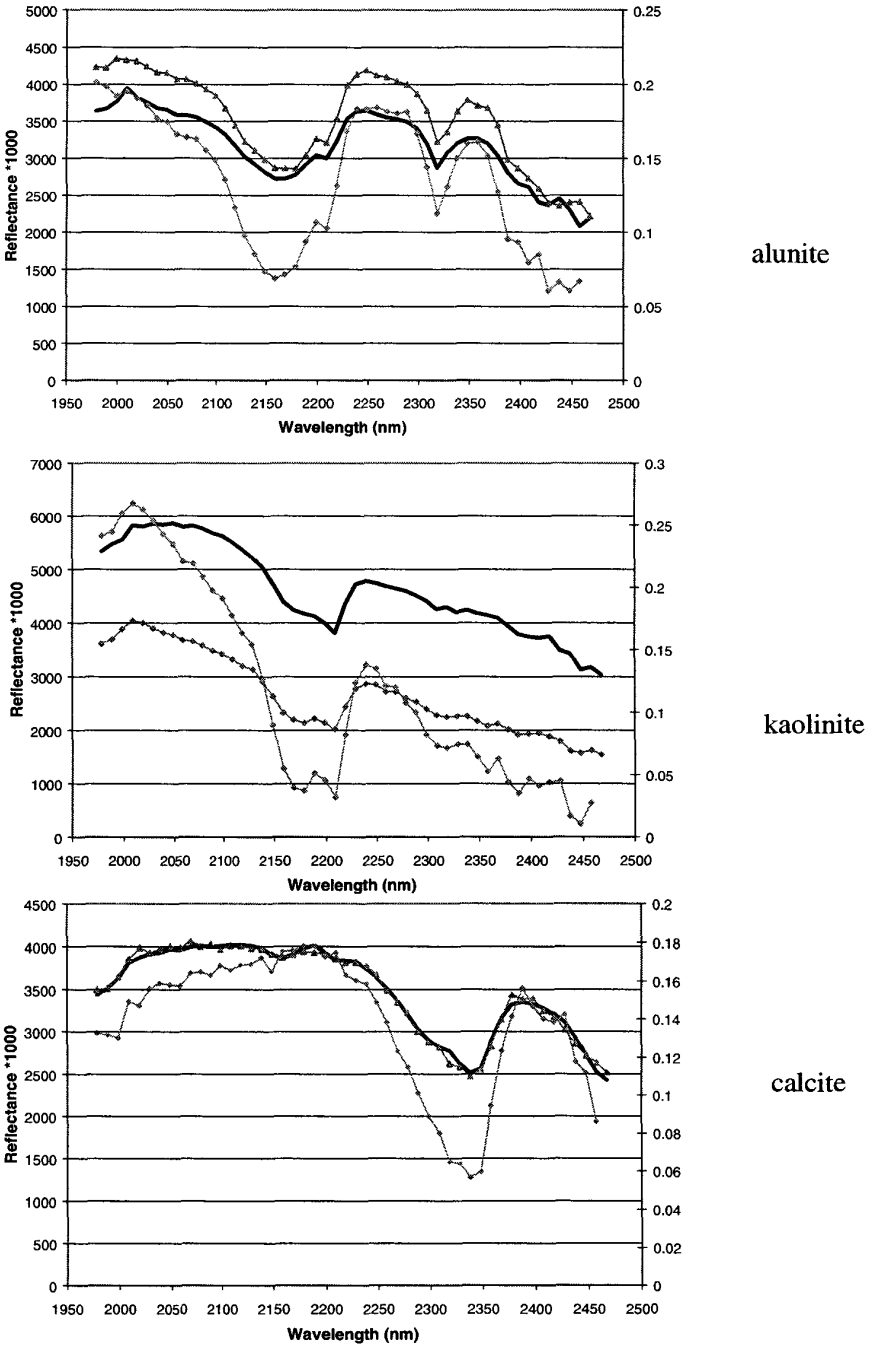


Figure 5. End-member Spectra for N-FINDR (solid black), Orasis (red) and SE (blue) Derived from Analyzing the Cuprite Data are identified with Alunite, Kaolinite and Calcite.

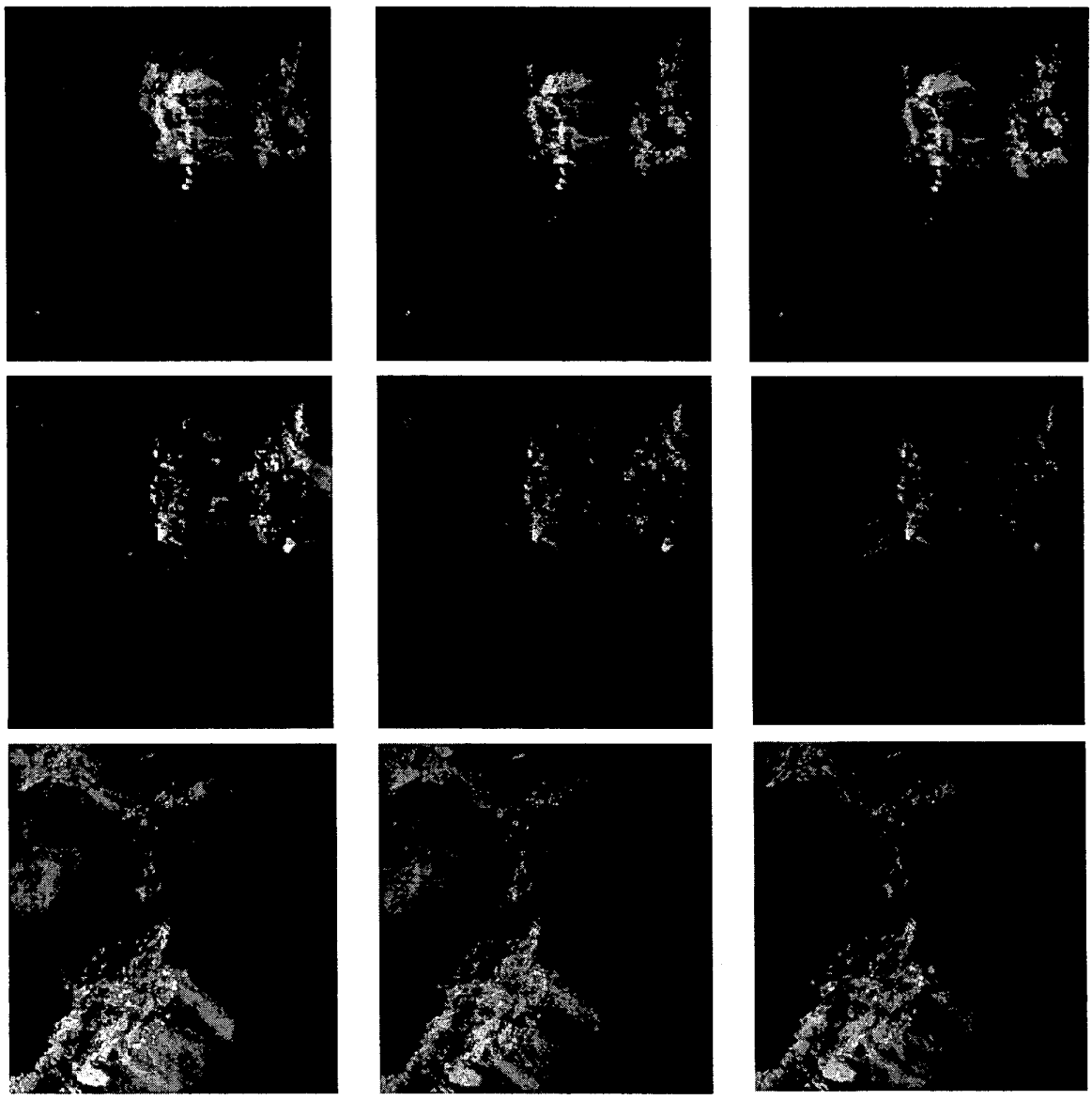


Figure 6. Fractional Abundance Planes for N-FINDR (left) and Orasis (center) and SE (right) Derived from Analyzing the Cuprite Data. Abundance Planes are identified with Alunite, Kaolinite and Calcite.

6. CONCLUSIONS

Several new tools are available for the analysis of the anticipated large volume of hyperspectral data expected from both airborne survey instruments and from satellites. Orasis has been applied successfully to a variety of scenes over the past several years. N-FINDR and IEA are new algorithms for the autonomous determination of end-members. These algorithms differ from each other and from Orasis in the method used to find end-members. N-FINDR is based on the concept of finding the maximum simplex that can be inscribed inside of a hyperspectral data set, while IEA uses an iterative unmixing approach. All three algorithms have been successfully used to determine end-members and unmix several hyperspectral data sets without the use of any *a priori* knowledge of the constituent spectra. In this paper, the derived end-members from Cuprite, Nevada AVIRIS images show a close match to USGS reference spectra [11], and the derived abundance maps are in good agreement with those published [12]. The comparison of the end-members and abundance planes produced by N-FINDR, Statistical Expansion (our implementation of the IEA), and Orasis shows that these autonomous algorithms produce similar results. At least, Orasis and N-FINDR are also capable of rapidly determining end-members and unmixing large scenes. The capability to perform this decomposition without use of *a priori* spectra is a powerful tool for hyperspectral analysis.

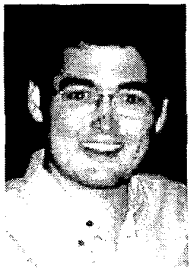
7. ACKNOWLEDGEMENTS

The authors wish to thank Michael Schlangen of Technical Research Associates, Inc. for his programming support and the reviewers for their helpful comments. In addition, the authors wish to acknowledge helpful discussions with Dr. Jeffrey Bowles of NRL and Dr. Scott Beaven of U. S. Navy SPAWAR Systems Center, San Diego.

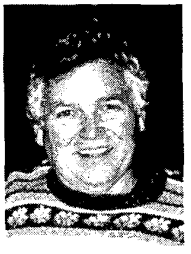
8. REFERENCES

1. Craig, M.D., "Minimum Volume Transforms for Remotely Sensed Data", IEEE Transactions on Geoscience and Remote Sensing, Vol. 32, pp. 542-552. 1994
2. Boardman, J, "Automating Spectral Unmixing of AVIRIS Data Using Convex Geometry Concepts", Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, 1, 11-14, JPL Pasadena. 1994
3. Bowles, J., P. Palmadesso, J. Antoniades, M. Baumbeck, and L.J. Rickard, "Use of Filter Vectors in Hyperspectral Data Analysis", Infrared Spaceborne Remote Sensing III, M. S. Scholl, B. F. Andresen, Editors, Proc. SPIE 2553, 148-157. 1995
4. 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, pp 337-344, Vancouver, B.C., Canada, 1999
5. Szeredi, T., K. Staenz, and R. Neville, "Automated Endmembers Selection: Part I Theory", Submitted for Publication in Remote Sensing of the Environment, 1999
6. Beaven, S., L. E. Hoff, and E. M. Winter "Comparison of SEM and Linear Unmixing Approaches for Classification of Spectral Data", to be published in SPIE Proceedings. 1999
7. Boardman, J. W., "Analysis, understanding and visualization of hyperspectral data as convex sets in n-space", SPIE Proceedings, Vol. 2480. pp. 14-22. 1995
8. Bowles, J., M. Daniel, J. Grossman, J. Antoniades, M. Baumback, and P. Palmadesso, "Comparison of Output from Orasis and Pixel Purity Calculations", Proceedings of SPIE, 3438, 148-156, 1998
9 Neville, R. A., Staenz, K, Szerdedi, T, Lebfebre, J. Hauff, P. "Automatic Endmember Extraction from Hyperspectral Data for Mineral Exploration", Proceeding of the Fourth International Airborne Remote Sensing Conference, Vol II pp 891-897, Ottawa, Ont. Canada, 1999
10 Vane, G., R. O. Green, T. G. Chrien, H. T. Enmark, E.G. Hansen, W. M. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," Remote Sensing of the Environment, Vol. 44, pp. 127-143, 1993
11 Clark, R.N., G.A. Swayze, A.J. Gallagher, T.V. King, and W.M. Calvin, The U. S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 microns, U.S. Geological Survey Open File Report 93-592, 1340 pages. 1993
12 Clark, Roger N. and Gregg A. Swayze, "Evolution in Imaging Spectroscopy Analysis and Sensor Signal-to-Noise: An Examination of How Far We Have Come", Summaries of the 6th Annual JPL Airborne Earth Science Workshop March 4-8, 1996 (available at http://speclab.cr.usgs.gov/PAPERS.imspec.evol/aviris.evolution.html)

9. BIOGRAPHIES



Michael E. Winter has been a consultant to Technical Research Associates, Inc. for three years. He received his B.S. in Geophysics from the University of California, Los Angeles in 1995 and is currently a candidate for a Ph.D in Geophysics at the University of Queensland, Brisbane Australia. He is the developer of the N-FINDR algorithm and has presented papers on Geophysics and Remote Sensing at American Geophysical Union meetings, SPIE and ERIM Environmental Conferences.



Edwin M. Winter is Research Director for sensor systems at Technical Research Associates, Inc. He received his Ph.D. in Geophysics from the University of California, Los Angeles in 1972 and has been active in remote sensing for twenty-eight years. He has presented papers at SPIE, IEEE, ERIM, EUROPTO, IRIS and other technical meetings and has published papers in the Physical Review, Remote Sensing of the Environment, Applied Optics as well as numerous conference proceedings.