"Precision Mining" of High-Dimensional Patterns with Self-Organizing Maps: Interpretation of Hyperspectral Images

Erzsébet Merényi

University of Arizona, Lunar and Planetary Laboratory, Tucson, AZ 85721, USA

Abstract. Utilization of remote sensing multi- and hyperspectral imagery has rapidly been increasing in numerous areas of economic and scientific significance. Hyperspectral sensors, in particular, provide the detailed spectral signatures that are known to uniquely characterize and identify minerals, soils, rocks, plants, aquatic species, man-made and other surface materials. This opens up tremendous possibilities for resource exploration and management, environmental monitoring, natural hazard prediction, and more. However, conventional analysis methods often prove inadequate for the exploitation of the wealth of information in hyperspectral images, because of the high dimensionality of the intricate spectra. Artificial Neural Networks hold the promise to revolutionize this area by meeting mathematical and other challenges that traditional techniques fail at. In this paper, we discuss the powers of Self-Organization for the detailed interpretation of hyperspectral images, using the full spectral dimensionality.

1 Remote Spectral Images as High-Dimensional Data

Airborne and satellite-borne spectral imaging has become one of the most advanced tools for collecting vital information about the surface covers of Earth and other planets. The utilization of these data includes areas such as mineral exploration, land use, forestry, ecosystem management; assessment of natural hazards, water resources, environmental contamination, biomass and productivity; and many other activities of economic significance, as well as prime scientific pursuits such as looking for possible sources of past or present life on other planets. The number of applications has dramatically increased in the past ten years with the advent of imaging spectrometers, that greatly surpass traditional multi-spectral sensors (*e.g.,* Landsat Thematic Mapper). Imaging spectrometers can resolve the known, unique, discriminating spectral features of minerals, soils, rocks, and vegetation. While a multi-spectral sensor samples a given wavelength window (typically the $0.4 - 2.5$ μm range in the case of Visible and Near-Infrared surface reflectance imaging) with several broad bandpasses, leaving large gaps between the bands, imaging spectrometers sample a spectral window contiguously with very narrow badpasses (Fig. 1). Hyperspectral technology is in great demand because direct identification of surface compounds is possible without prior field work, for materials with known spectral signatures.

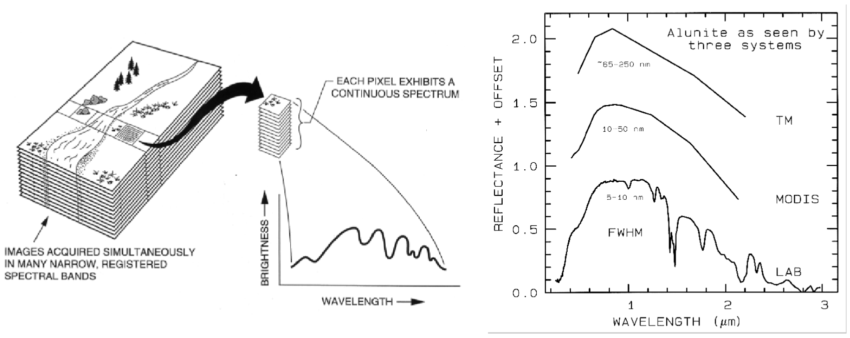

Fig. 1. Left: The concept of hyperspectral imaging. Figure from [1]. Right: The spectral signature of the mineral alunite as seen through the 6 broad bands of Landsat TM, as seen by the moderate spectral resolution sensor MODIS (20 bands in this region), and as measured in laboratory. Hyperspectral sensors such as AVIRIS of NASA/JPL [2] produce spectral details comparable to laboratory measurements.

Formally, the vector $S^{x,y} = (S_1^{x,y}, ..., S_{NB}^{x,y})$, where $S_k^{x,y}$ is the data value in the kth image band ($k = 1, ..., NB$) at pixel location (x, y), is called a spectrum. It is a characteristic pattern (Fig. 1) that provides a clue to the surface material(s) within pixel (x, y). NB denotes the number of image bands. $x = 1, ..., Xmax$ and $y = 1, ..., Ymax$, where $Xmax$ and $Ymax$ are the spatial size of the image in the x and y direction, respectively. The feature space spanned by Visible-Near Infrared reflectance spectra is $[0 - noise, U + noise]^{NB} \subset \Re^{NB}$ where $U > 0$ represents an upper limit of the measured scaled reflectivity and $noise$ is the maximum value of noise across all spectral channels and image pixels. Sections of this space can be very densely populated while other parts may be extremely sparse, depending on the materials in the scene and on the spectral bandpasses of the sensor.

Great spectral detail comes at a cost of very high data volume and complexity. It poses new mathematical challenges in the clustering and classification of images with high spectral dimensionality. The specific problems associated with remote sensing spectral image analyses arise from any combination of the following:

- The spectral patterns are high dimensional (dozens $\leq NB \leq$ hundreds);
- The number of data points (image pixels) can be as large as several millions;
- Surface materials that are significantly different for an application may be distinguished by very subtle differences in their spectral patterns;
- Given the richness of data, the goal is to separate many cover classes, or to separate materials with subtle spectral differences;
- The pixels can be mixed: Several different materials may contribute to the spectral signature detected from each pixel;
- Very little training data may be available for some classes; and classes may be represented very unevenly.

Noise is far less problematic than the intricacy of the actual spectral patterns, because of the high Signal-to-Noise Ratios (500 – 1,500) that present-day hyperspectral imagers provide. For this discussion, we will omit noise issues, and additional effects such as atmospheric distortions, illumination geometry and albedo variations in the scene, because these can be addressed through well-established procedures prior to clustering or classification.

ANNs have been gaining recognition as powerful answers to the above challenges. Many sucessful applications were published, that analyze lower dimensional spectral data (such as Landsat TM). See, for example, [3], and references therein. Few have attempted ANN (or any) classification of hyperspectral data into a large number of classes even though that is where the shortcomings of classical methods become a severe limiting factor in accomplishing scientific or technical objectives. For hyperspectral images dimensionality reduction is frequently accepted in order to accomodate the data to the capability of traditional methods. This, however, often results in undesirable loss of information, preventing "precision mining": discovery of new, interesting classes, or discrimination of classes with subtle spectral differences. Thus the analysis negates the purpose of the sophisticated sensors.

The rest of the paper is structured as follows: In Section 2, the test data set, used throughout this paper, is introduced. Section 3 discusses hyperspectral dimensionality. Section 4 and 5 present results of clustering and classification of the test data, using the full hyperspectral dimensionality; and in Setion 6, correctness and relaiabilty issues are discussed.

2 Test Data Set: The Lunar Crater Volcanic Field AVIRIS Image

In this paper a VIS-NIR (0.4 - 2.5 μm), 224-band, 30 m/pixel AVIRIS image of the Lunar Crater Volcanic Field (LCVF), Nevada, USA, is used. The LCVF is one of NASA's remote sensing test sites, where comprehensive field studies were conducted [4]. In addition, much accumulated field knowledge and independent geologic research by individuals provide excellent and detailed basis for evaluation of the results presented here [5].

Fig. 2 shows a natural color composite of the LCVF with locations representative of 23 cover classes marked by their respective class labels. This 512 x 614 pixel AVIRIS image of a 10 x 12 square km area contains, among other materials, volcanic cinder cones (class A, reddest peaks) and weathered derivatives thereof such as ferric oxide rich soils (L, M, W), basalt flows of various ages (F, G, I), a dry lake divided into two halves of sandy (D) and clayey composition (E); a small rhyolitic outcrop (B); and some vegetation at the lower left corner (J), and along washes (C). Alluvial material (H), dry (N,O,P,U) and wet (Q,R,S,T) playa outwash with sediments of various clay

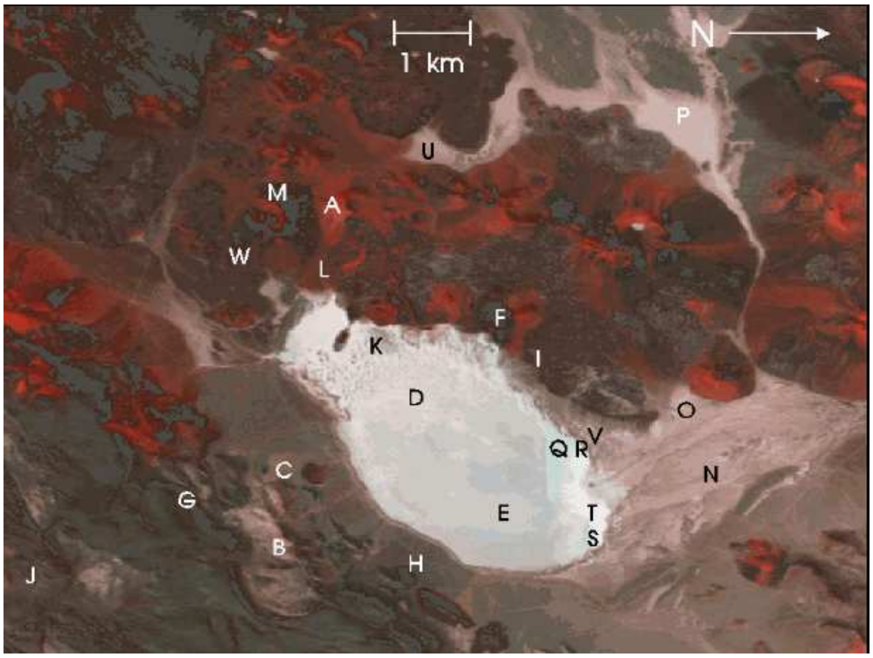

Fig. 2. The Lunar Crater Volcanic Field. Color composite from an AVIRIS, 1994 image. The original image comprises 224 image bands over the 0.4 - 2.5 μm range, 512 x 614 pixels, altogether 140 Mbytes of data. Labels indicate different cover types described in the text. The ground resolution is 30 m/pixel.

content as well as other sediments (V) in depressions of the mountain slopes, and basalt cobble stones strewn around the playa (K) form a challenging series of spectral signatures for pattern recognition (see in [6]). A long, NW-SE trending scarp, straddled by the label G, borders the vegetated area. On this color composite, containing information from only three image bands, many of the cover type variations do not manifest. They will become visible in the cluster and class maps below. The image comprises 140 Mb of data. After atmospheric correction and removal of excessively noisy bands (saturated water bands and overlapping detector channels), 194 image bands remained. These 194-dimensional spectra are the input patterns in the following analyses.

3 The Dimensionality of Hyperspectral Patterns

The spectral dimensionality of hyperspectral images is not well understood. Many believe that hyperspectral images are highly redundant because of band correlations. Others maintain an opposite view, which also manifests in the vigorous development of hyperspectral sensors and commercialization of hyperspectral data services in the last several years. Few investigations exist yet into the intrinsic dimensionality (ID) of hyperspectral images. Linear methods such as PCA or determination of mixture model endmembers using either of the two major approaches [7] [8] usually yield 3–8 "effective components". Bruske [9] finds the spectral ID of the LCVF AVIRIS image (Fig. 2) to be between 4 and 7, using a non-linear Neural Net based approach, Optimally Topology Preserving Maps. These surprisingly low numbers, that increase with improved sensor performance [10], result from using statistical thresholds for the determination of what is "relevant", regardless of application dependent criteria.

The number of relevant components increases dramatically when specific goals are considered such as what cover classes should be separated or what known properties of the surface can be postulated. Pendock [11], using an associative neural network, extracted 20 linear mixing endmembers from a 50-band (2.0-2.5 μm) segment of an AVIRIS image (Fig. 2), requiring only that a rather general surface texture criterium be met. Benediktsson et al. [12] performed feature extraction on an AVIRIS geologic scene, which resulted in 35 bands. They used an ANN (the same network that performed the classification itself) for Decision Boundary Feature Extraction (DBFE). The DBFE is claimed to preserve all features that are necessary to achieve the same accuracy as in the original data space, by a given classifier for predetermined classes. However, no comparison of classification accuracy was made using the full spectral dimension, to support the DBFE claim. In this particular study a relatively low number of classes, 9, were of interest, and the question posed was to find the number of features to describe those classes. Separation of a higher number of classes may require more features.

It is not clear how feature extraction should be done in order to preserve relevant information in hyperspectral images. Later in this paper it is demonstrated that extraction of about 30 features from our LCVF image by any of several methods leads to a loss of a number of the originally determined 23 cover classes. Wavelet compression studies on an earlier image of the the same AVIRIS scene [13] conclude that various schemes and compression rates affect different spectral classes differently, and none was found overall better than another, within 25% – 50% compressions (retaining 75% – 50% of the wavelet coefficients). In a study on simulated, 201-band spectral data, [14] show slight accuracy increase across classifications on 20-band, 40-band, and 60-band subsets. However, they base the study on only two vegetation classes, the feature extraction is a progressive hierarchical subsampling of the spectral bands, and there is no comparison with using the full, 201-band case. Comparative studies using full spectral resolution and many classes are lacking, in general, because few methods can cope with such high-dimensional data technically, and the ones that are capable (such as Minimum Distance, Parallel Piped) often perform too poorly to merit consideration.

Undesirable loss of relevant information can result using any of these feature extraction approaches. In any case, finding an optimal feature extraction requires great preprocessing efforts just to taylor the data to available tools. An alternative is to develop capabilities to handle the full spectral information. Analysis of unreduced data is important for exploration and novelty detection (such as in the case of hard-earned data in planetary exploration); as well as to allow for the distinction of significantly greater number of cover types, according to the potential provided by modern imaging spectrometers.

For example, remote monitoring of the Colorado River ecosystem in Grand Canyon, Arizona, USA, requires discrimination and identification (spatial mapping) of about seventy plant and aquatic species, in order to provide a meaningful information service for 26 stakeholders in businesses of dam operations, electric power generation, recreation, tourism, cultural affairs of native American tribes, and scientific research. Some of the species exhibit very slight but systematic spectral differences that are captured in hyperspectral data. These differences would most probably be lost in a PCA or a wavelet transformation, and possibly in more sophisticated algorithms. Preliminary SOM-based classification, utilizing the full spectral information, was successful in separating several "doubtful" cover types [15]. This example represents a modern emerging paradigm that is based on the expectation of continuous availability of (satellite-based) hyperspectral imagery and that the information content of such data, if extracted properly, can satisfy a wide range of needs simultaneously.

4 Effects of Dimensionality Reduction and Inadequate Tools

A systematic study was conducted on the LCVF image (Fig. 2), to simultaneously assess loss of information due to reduction of spectral dimensionality, and to compare performances of several traditional and an SOM-based hybrid ANN classifier. 23 geologically relevant classes, as indicated in Fig. 2, were used. These represent a great variety in terms of spatial extent, the similarity of spectral signatures [6], and the number of available training samples. The full study is described in [16]. Here the most dramatic finding is presented.

Fig. 3, top panel, shows the best classification, produced by an SOM-hybrid ANN using all 194 spectral bands. This ANN first learns in an unsupervised mode, during which the input data are clustered in the hidden SOM layer. After the SOM converges, the output layer is allowed to learn class labels. The preformed clusters in the SOM greatly aid in accurate and sensitive classification, by helping prevent the learning of inconsistent class labels. Detailed description of this classifier is given in several previous scientific studies, which produced improved interpretation of high-dimensional spectral data compared to earlier analyses [17] [18] [19]. Training samples for the supervised classifications in our study were selected based on knowledge of and interest in geologic units. The SOM hidden layer was not evaluated and used for identification of spectral types, prior to training sample determination. So, Fig. 3 reflects the geologist's view of the desirable segmentation.

In order to apply Maximum Likelihood and other covariance based classifiers, the number of spectral channels needed to be reduced to 30, since the maximum number of training spectra that could be identified for *all* classes was 31. Dimensionality reduction was performed in several ways, including PCA, equidistant subsampling, and band selection by a domain expert. Band selection by domain expert proved most favorable. Fig. 3, bottom panel, shows the Maximum Likelihood classification on the LCVF data, reduced to 30 bands. A number of classes (notably the ones with subtle spectral differences, such as N, Q, R, S, T, V, W) were entirely lost. Class K (basalt cobbles) disappeared from most of the edge of the playa, and only traces of B (rhyolitic outcrop) remained. Class G and F were greatly overestimated. Although the ANN classifier produced better results (not shown here) on the same 30-band reduced data set than the Maximum Likelihood, a marked drop in accuracy occurred compared to classification on the full data set. This emphasizes that accurate mapping of "interesting", spatially small geologic units is possible from full hyperspectral information and with appropriate tools.

It should be noted that a Backpropagation (BP) network is very difficult to train for the combination of such high number of inputs and output classes. This author has not seen hyperspectral images classified by a BP network, at full spectral resolution. In contrast, the SOM-hybrid ANN above scales up gracefully and performs the task with relative ease. Inclusion of unsupervised learning phases was recently shown to increase the classification accuracy of a BP network too [20].

5 Discovery in Hyperspectral Images with SOMs

Let's examine now how the SOM preforms in terms of detection of clusters in high-dimensional data. Fig. 4 displays a 40 x 40 SOM that was generated by a modification of the original SOM [21], the conscience algorithm of DeSieno [22]. The input data space was the entire 194-band LCVF image. Groups of neurons, altogether 32, that were found to be sensitized to groups of similar spectra in the 194-dimensional LCVF input data, are indicated by various colors. The boundaries of these clusters were determined by a somewhat modified version of the Ultsch and Simeon method [23]. Areas where no data points (spectra) were mapped are the grey corners with uniformly high fences, and are relatively small. The black background in the SOM lattice shows areas that have not been evaluated for cluster detection. The spatial locations of the image pixels mapped onto the groups of neurons in Fig. 4, are shown in the same colors in Fig. 5. Color coding for clusters that correspond to classes or subclasses of those in Fig. 3, top, is the same as in Fig. 3, to show similarities. Colors for additional groups were added.

The first observation is the striking match between the supervised ANN class map in Fig. 3, top panel, and this clustering: the SOM detected all classes that were known to us as meaningful geological units. The "discovery" of classes B (rhyolitic outcrop, white), F (young basalt flows, dark grey and black, some shown in the black ovals), G (a different basalt, exposed along the scarp, dark blue, one segment outlined in the white rectangle), K (basalt cobbles, light blue, one segment shown in the black rectangle), and other spatially small classes such as the series of playa deposits (N, O, P, Q, R, S, T) is significant. This is the capability we need for sifting through high-volume, high-information-content data to be alerted to interesting, novel, or hard-to-find units. The second observation is that the SOM detected more, spatially coherent, clusters than the number of classes that we trained for in Fig. 3. The SOM's view of the data is more refined and more precise than that of the geologist's. For example, class A (red in Fig. 3) is split here into a red (peak of cinder cones) and a dark orange (flanks of cinder cones) cluster, that make geologic sense. The maroon cluster to the right of the red and dark orange clusters at the bottom of the SOM fills in some areas that remained unclassified (bg) in the ANN class map, in Fig. 3. An example is the arcuate feature at the base of the cinder cone in the white oval, that apparently contains a material different enough to merit a separate spectral cluster. This material fills other areas too, consistently at the foot of cinder cones (another example is seen in the large black oval). Observation of further refinements are left to the reader.

6 Correctness, Reliability, and Precision of SOM Analyses

The correctness of the SOM mapping is a serious concern. For small data sets such as in [17], and for a low number of clusters, topology violation can be detected by semi-manual examination of the map, and the situation remedied by using a larger lattice or often by just retraining the SOM starting from a different random initial state. For large data sets, the effort needed to detect the ill conditions are great, and repeated runs are expensive. Hyperspectral dimension is of particular concern because very few SOM studies exist on such high-dimensional data, to draw experience from [24]. A measure that expresses the level of topology violation in mapping an n-dimensional data space onto lower dimension has been constructed by Villmann, Bauer et al. [25] [26]. This work is based on findings by Martinetz and Schulten [27], and carry further in that it provides a formula that can be computed from the weights and neighborhood relations, and immediately alert for entanglement of clusters. The work of Villman et al. also provides relief from topology violation by automatically growing the SOM lattice [28] [29]. Applications to 6-band Landsat images worked well [30]. Further work is underway to evaluate the powers of this method for hyperspectral imagery [31].

A mapping that matches the probability density (*pdf*) of the input data is preferable, because it represents regions of the input space with resolutions appropriate for their relative densities. For this, the map magnification factor $m(x)$, which is the number of neurons representing a subsection dx of the input space, is required to be proportional to the *pdf*. The original SOM algorithm does not meet the above requirement and tends to overrepresent regions of low density, and underrepresent regions of high density [32]. This could prevent sensitive separation of spectral clusters with subtle signature differences, such as the sandy and clayey part of the playa (classes D and E), or the splitting of class A into the red and dark orange clusters (Fig. 3, bottom panel and Fig. 5). The *conscience algorithm* of DeSieno [22] adjusts the winning frequencies of the SOM neurons so as to ensure proper spreading (compaction) of information that is dense (sparse) in the input feature space. Update of the weight vectors is given by $w_i^{t+1} = w_i^t + \alpha^t(S - w_i^t)$, for $i \in N_c$ and $w_i^{t+1} = w_i^t$ otherwise, where c and N_c denote, respectively, the index and the the index set of the neighbourhood of the winning neuron PE_c. $c = arg\min_i(\|S - w_i\| - B_i)$ where the bias term $B_i = \gamma(1/M - F_i^t)$ adjusts the distance between the incoming pattern S and w_i based on the historic winning frequency F_i of PE_i. M is the number of PEs in the Kohonen lattice. F_i is updated along with the weights, according to $F_i^{t+1} = F_i^t + \beta(\delta_{N_c} - F_i^t)$ where δ_{N_c} is 1 for $i \in N_c$, 0 otherwise. α, β and γ are user controlled parameters decreasing in time. In his new book, van Hulle [33] points out that adding a conscience algorithm to the SOM does not equate to equiprobabilistic mapping, in general. However, for *very high dimensions*, a minimum distortion quantizer (such as the conscience algorithm) approaches an equiprobable quantizer [33] (page 93). Fig. 4, 5, and 6 seem to support this statement, although precise evaluation is difficult because of the long time needed to reach a mature state of the SOM for large hyperspectral images.

Visualization of the SOM's knowledge is also crucial for "precision mining". Detection of clusters — as a post-processing step after SOM convergence — hinges on how well the SOM's internal view of the data space is communicated through knowledge extraction and representation. This has been targeted by several works [23] [34] [35] [36] [37]. However, ours is the first application to hyperspectral images. Closest in spirit to the browsing capabilities of [37] and the WEBSOM as used in [24], our applications require entirely different engineering solutions because of the specifics of spectral-spatial image data. In our facilities, developed at the Lunar and Planetary Laboratory, University of Arizona, interactive cluster determination tools include "knobs" for on-the-fly adjustment of fence scaling and thresholding in the Ultsch-Simeon representation; changing the underlying map from "plain" (as in Fig. 4) to "density map" (where the number of data points mapped on a grid cell is indicated by grey scale or monochrome color intensities); switching between separate or combined viewing of map and fences; changing colors, and size of grid cells and fences, for clearest possible viewing, etc. Different representations such as [35] are also planned. The ultimate goal is automation of cluster extraction. Alternative approaches such as [38] are being evaluated.

7 Acknowledgements

Software develpment for this work has been supported by the Applied Information Systems Research Program of the Office of Space Science, NASA, NAG54001. NeuralWare and Khoros packages were utilized. I am grateful to Dr. Thomas Villmann for valuable discussions on various aspects of SOMs. Many thanks are due to Dr. William H. Farrand for providing field and geologic knowledge in evaluating results for the LCVF; and to many interdisciplinary collaborators for their participation in the large-scale, real-life applications that provide the test bed for this research.

References

1. Campbell, J. (1996) Introduction to Remote Sensing. The Guilford Press
2. Green, R. O. Ed. (1996) Summaries of the 6th Annual JPL Airborne Geoscience Workshop, **1** AVIRIS Workshop, Pasadena, CA, March 4–6, 1996
3. Carpenter, G. A., Gjaja, M. N., et al. (1997) ART Neural Networks for Remote Sensing: Vegetation Classification from Landsat TM and Terrain Data, IEEE. Trans. Geosci. and Remote Sens. **35**(2), 308–325
4. Arvidson, R. E., Dale-Bannister, M. et al. (1991) Archiving and distribution of Geologic Remote Sensing Field Experiment Data. EOS, Transactions of the American Geophysical Union, **72**(17), 176
5. Farrand, W. H. (1991) VIS/NIR Reflectance Spectroscopy of Tuff Rings and Tuff Cones. Ph.D. Thesis, University of Arizona
6. Merényi, E., (1998) Self-Organizing ANNs for Planetary Surface Composition Research, Proc. 6th European Symposium on Artificial Neural Networks, ESANN'98, Bruges, Belgium, April 22–24, 1998, 197–202
7. Adams, J. B., Smith, M. O., and Gillespie, A. R. (1993) Imaging spectroscopy: Interpretation based on spectral mixture analysis, In Remote Geochemical Analysis: Elemental and Mineralogical Composition. C.M. Peters and P.A.J. Englert, Eds., Cambridge University Press, New York, 145–166
8. Research Systems Inc. (1997) ENVI v.3 User's Guide, RSI, Boulder, 614 pp.
9. Bruske, J., and Merényi, E. (1999) Estimating the Intrinsic Dimensionality of Hyperspectral Images. Proc. 7th European Symposium on Artificial Neural Networks, ESANN'98, Bruges, Belgium, April 21–23, 1999, 105–110

10. Green, R. O. and Boardman, J. (2000) Exploration of the relationship between Information content and Signal-to-Noise ratio and spatial resolution. Proc. 9th AVIRIS Earth Science and Applications Workshop, February 23–25, Pasadena, CA (in print).
11. Pendock, N. (1999) A Simple Associative Neural Network for Producing Spatially Homogeneous Spectral Abundance Interpretations of Hyperspectral Imagery. Proc. 7th European Symposium on Artificial Neural Networks, ESANN'98, Bruges, Belgium, April 21–23, 1999, 99–104
12. Benediktsson, J. A., Sveinsson, J. R., et al. (1994) Classification of Very-High-Dimensional Data with Geological Applications, Proc. MAC Europe 91, Lenggreis, Germany, 4–6 October, 1994, 13–18
13. Moon, T., and Merényi, E. (1995) Classification of hyperspectral images using wavelet transforms and neural networks. Proc. of the Annual SPIE Conference, July 9–14, San Diego, CA **2569**
14. Benediktsson, J. A., Swain, P. H., et al. (1990) Classification of Very High Dimensional Data Using Neural Networks, IGARSS'90 10th Annual International Geoscience and Remote Sensing Symp. **2**, 1269
15. Merényi, E., Farrand, W. H. et al. (2000) Studying the Potential for Monitoring Colorado River Ecosystem Resources Below Glen Canyon Dam Using Low-Altitude AVIRIS Data. Proc. 9th AVIRIS Earth Science and Applications Workshop, February 23–25, Pasadena, CA (in print).
16. Merényi, E., Farrand, W. H. et al. (2000) Efficient Geologic Mapping from Hyperspectral Images with Artificial Neural Networks Classification: a Comparison to Conventional Tools. In preparation for submission to IEEE TGARS
17. Howell, E. S., Merényi, E., Lebofsky, L. A. (1994) Classification of Asteroid Spectra Using a Neural Network Jour. Geophys. Res. **99**, 10,847–10,865
18. Merényi, E., Howell, E. S., et al. (1997) Prediction of Water In Asteroids from Spectral Data Shortward of 3 Microns. ICARUS **129**, 421–439
19. Merényi, E., Singer, R. B., Miller, J. S. (1996) Mapping of Spectral Variations On the Surface of Mars From High Spectral Resolution Telescopic Images, ICARUS **124**, 280–295
20. Fardanesh, M. T., and Ersoy, O. K. (1998) Classification Accuracy Improvement of Neural Network Classifiers by Using Unlabeled Data. IEEE Trans. Geosci. and Remote Sens.**36**(3), 1020–1025
21. Kohonen, T. (1997) Self-Organizing Maps. Springer Series in Information Sciences, **30**, Springer, Berlin, Heidelberg, New York, 1995, 1997
22. DeSieno, D. (1988) Adding a Conscience to Competitive Learning. Proc. ICNN, New York, July 1988 **I**, 117–124
23. Ultsch, A. and Simeon, H. P. (1990) Kohonen's Self Organizing Feature Map for Exploratory Data Analysis. Proc. INNC-90-PARIS **I**, 305–308
24. Kohonen, T. (1997) Exploration of Very Large Databases by Self-Organizing Maps. Proc. IEEE ICNN'97 **I**, PL1–6
25. Bauer, H.U., Herrmann, M., Villmann, Th. (1997) Topology Preservation in Neural Maps of Real World Data. Proc. 5th European Symposium on Artificial Neural Networks, ESANN'97, Bruges, Belgium, April 22–24, 1997, 205–210
26. Villmann, Th., Herrmann, R.Der, and Martinetz, Th. (1997) Topology Preservation in Self-Organizing Feature Maps: Exact Definition and Measurement. IEEE Trans. on Neural Networks **8**(2), 256–266
27. Martinetz, Th., and Schulten, K. (1994) Topology Representing Networks. Neural Networks **7**(3), 507–522

28. Bauer, H.U. and Villmann, Th. (1997) Growing a Hypercubical Output Space in a Self-Organizing Feature Map. IEEE Trans. on Neural Networks **8**(2), 218–226
29. Villmann, Th. and Bauer, H.-U. (1998) Applications of the growing self-organizing map. Neurocomputing **21**, 91–100
30. Villmann, T. (1999) Benefits and Limits of the Self-Organizing Map and its Variants in the Area of Satellite Remote Sensoring Processing. Proc. 7th European Symposium on Artificial Neural Networks, ESANN'98, Bruges, Belgium, April 21–23, 1999, 111–116
31. Villmann, T. and Merényi, E. (2000) Extensions and Modifications of SOM and its Application in Satellite Remote Sensing Processing. Accepted to Proc. 2nd Int'l Computer Science Conventions Symposium on Neural Computation, NC'2000, May 23–26, 2000, Berlin, Germany
Conference on Geologic Remote Sensing, San Antonio, TX, 9–12 May,
32. Haykin, S. (1995) Neural Networks. A Comprehensive Foundation. McMillan
33. van Hulle, M. M. (2000) Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organization. Wiley & Sons
34. Mao, J. and Jain, A. K. (1995) Artificial Neural Networks for Feature Extraction and Multivariate Data Projection. IEEE Trans. on Neural Networks **6**(2), 296–317
35. Cottrell, M. and de Bodt, E. (1996) A Kohonen Map Representation to Avoid Misleading Interpretations. Proc. European Symposium on Artificial Neural Networks, Bruges, Belgium, 22–24 April, 1996
36. Kraaijveld, M. A., Mao, J. and Jain, A. K. (1995) A Nonlinear Projection Method Based on Kohonen's Topology Preserving Maps. IEEE Trans. on Neural Networks **6**(3), 548–559
37. Kaski, S., Nikkila, J. and Kohonen, T. (1998) Methods for Interpreting a Self-Organizing Map in Data Analysis. Proc. 6th European Symposium on Artificial Neural Networks, ESANN'98, Bruges, Belgium, April 22–24, 1998, 185–190
38. Fyfe, C. and Baddeley, R. (1995) Non-linear data structure extraction using simple hebbian networks. Biol. Cybern. **72**, 533–541

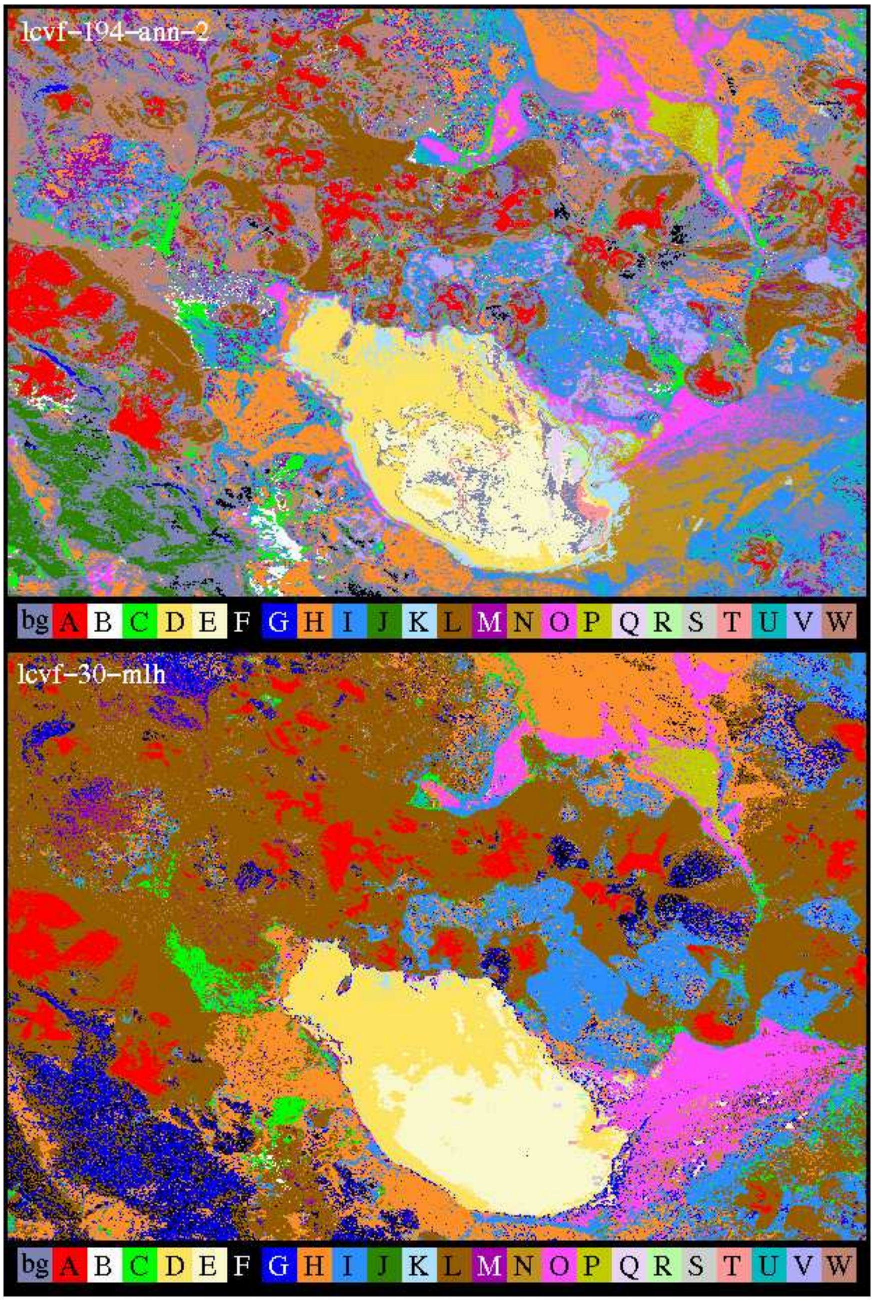

Fig. 3. Top: SOM-hybrid ANN classification of the LCVF scene, using 194 image bands. Bottom: Maximum Likelihood classification of the LCVF scene. 30, strategically selected bands were used due to the limited number of training samples. Considerable loss of class distinction occurred compared to the ANN classification. 'bg' stands for background (unclassified pixels).

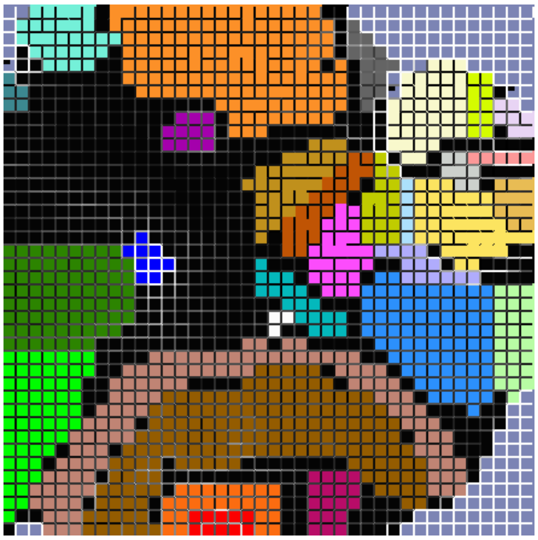

Fig. 4. Clusters identified in a 40 x 40 SOM. The SOM was trained on the entire 194-band LCVF image, using the DeSieno [22] algorithm.

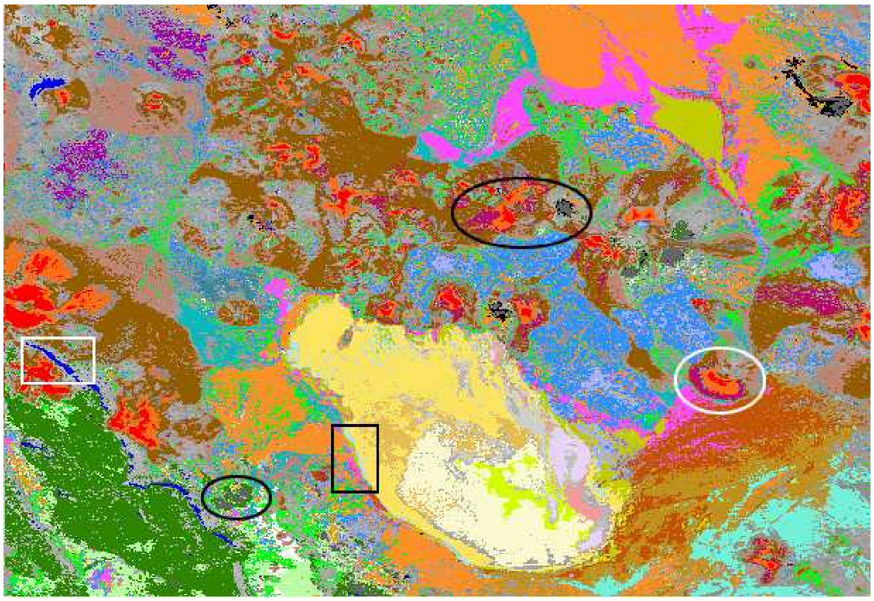

Fig. 5. The clusters from Fig. 4 remapped to the original spatial image, to show where the different spectral types originated from. The relatively large, light grey areas correspond to the black, unevaluated parts of the SOM in Fig. 4. Ovals and rectangles highlight examples discussed in the text.

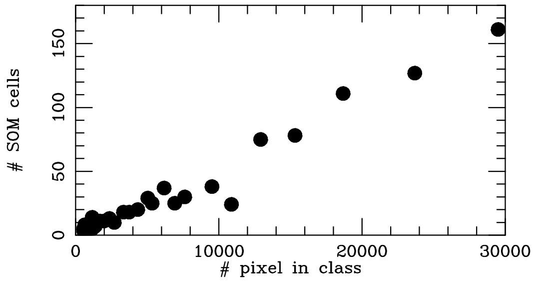

Fig. 6. Relationship between the number of image pixels in the clusters in Fig. 5, and the number of SOM neurons representing them in Fig. 4.