Neural Maps for Precision Data Mining: Application to Planetary Spectral Images

Erzsébet Merényi
Electrical and Computer Engineering, Rice University MS 380
Houston, TX 77005, U.S.A.
e-mail: erzsebet@rice.edu

Abstract

The use of neural maps, specifically Self-Organizing Maps, is shown for "precision" mining of high-dimensional scientific data. The motivation comes from the computational challenges posed by complex multiband spectral images acquired by advanced remote sensors. Today spectral and hyperspectral imagers are present on virtually every Earth-orbiting satellite and space mission, acquiring immense amount of very rich data. Conventional clustering and classification methods often meet their theoretical or practical limitations when confronted with such data. This paper reviews the mathematical challenges, the limitations of some classical favorite methods, neural maps, and presents sophisticated material identification from spectral images with advanced Self-Organizing Maps.

Keywords: Artificial Neural Networks, Self-Organizing Maps, hyperspectral imaging, data mining, classification

I. THE INFORMATION IN SPECTRAL IMAGES AND THE CHALLENGES OF EXTRACTING IT

Airborne and satellite-borne spectral imaging has become one of the most important tools for collecting vital information about the surface covers of Earth and other planets. The utilization of these data includes mineral exploration, land use, forestry, ecosystem management; assessment of natural hazards, water resources, environmental contamination, biomass and productivity; and many other areas of economic and scientific significance, such as looking for possible signs of past or present life on other planets. The data acquired by spectral imagers are "stacked" images of the same spatial area, each taken at a different wavelength. The individual images are called image bands (Fig. 1).[1] *Hyperspectral* sensors, developed in the past 10–15 years, acquire as many as 100–500 image bands simultaneously, contiguously covering a given window of the electromagnetic spectrum at very small wavelength increments. 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*. Since materials reflect incident sunlight preferentially at various wavelengths, the resulting spectrum is a characteristic, repeatable pattern which provides a unique identification of the surface material(s) within pixel (x, y). Surface reflectance spectroscopy uses the visible and near infrared (VIS-NIR) wavelength range, typically from 0.4 to 2.5 μm. The feature space spanned by VIS-NIR spectra is $[0, U]^{NB} \subset \Re^{NB}$ where $U > 0$ represents an upper limit of the measured scaled reflectivity. 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 resolution of the sensor.

Examples of hyperspectral imagers are the Airborne Visible-Near-Infrared Imaging Spectrometer (AVIRIS, NASA/JPL [1]), Hydice (Naval Research Lab, [2]), SpecTIR (http://www.spectir.com/), Hyperion on the EO-1 Earth Observing satellite (http://eo1.gsfc.nasa.gov/Technology/Hyperion.html), CASI (ITRES, Canada, http://www.itres.com/), the Visible-Infrared Mapping Spectrometer (VIMS) on the Cassini spacecraft, orbiting Saturn since June, 2004, (http://saturn.jpl.nasa.gov/spacecraft/instruments-cassini-intro.cfm). See a longer list at http://www.geo.unizh.ch/~schaep/research/apex/is_list.html . Detailed clustering and classification of hyperspectral imagery can provide a great wealth of information for scientists and decision makers. However, the intricate spectral patterns pose unique challenges for any of the following reasons:

- The patterns are high dimensional (dozens $\leq NB \leq$ hundreds);
- The number of data points (image pixels) can be very large, $\geq$ 1M;
- Given the richness of data, the goal is to separate many cover classes;

[1] Original figures are in color. Color paper posted at http://www.ece.rice.edu/~erzsebet/publications.html

• Different materials may be distinguished by only subtle differences in their spectral patterns;

• Very little training data may be available for some classes; and classes may be represented very unevenly.

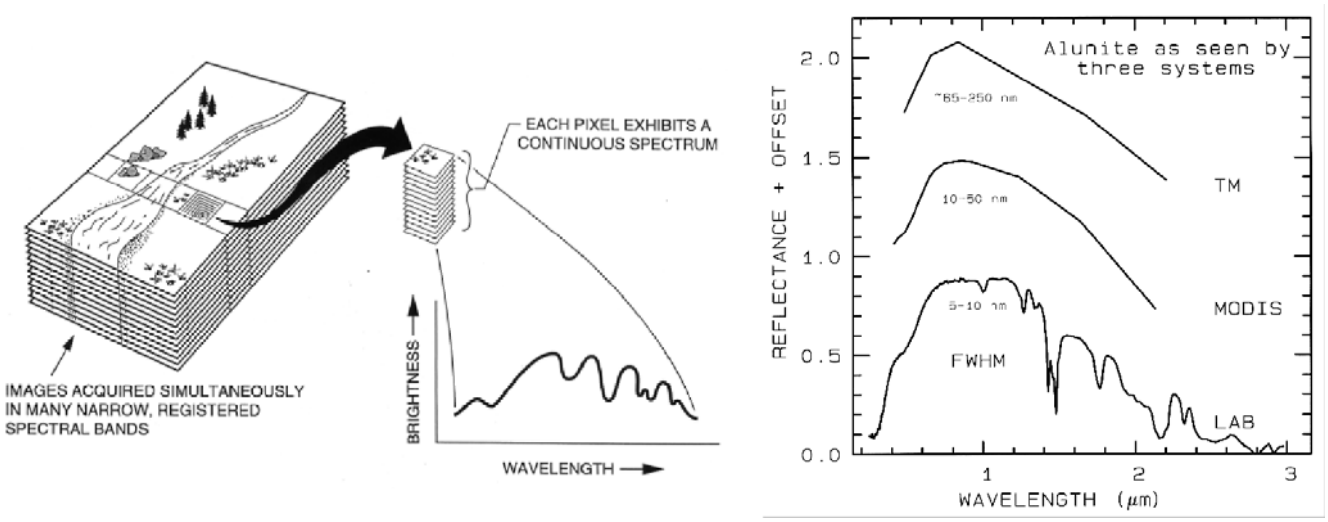

Fig. 1. **Left:** The concept of hyperspectral imaging. Figure from [3]. **Right:** The spectral signature of the clay mineral alunite as seen through the 6 broad bands of Landsat TM, as seen by the moderate spectral resolution sensor MODIS (20 bands in this region), and as measured in laboratory. Figure from [4]. Hyperspectral sensors such as AVIRIS of NASA/JPL [1] produce spectral details comparable to laboratory measurements.

Additional complications arise from atmospheric distortions, noise, illumination geometry and albedo variations in the scene. These can be treated with established preprocessing methods prior to classification, and therefore are not discussed here.

Favorite traditional classifiers and clustering algorithms (such as Maximum Likelihood, Parallel Piped, Mahalonobis Distance, K-Means or Isodata) have difficulty handling many image bands (the high-dimensional data vectors $S^{x,y}$ as input patterns) and therefore perform poorly. For example, the Maximum Likelihood and other covariance based classifiers require at least $NB + 1$ training samples for *each* class in order to avoid a singular covariance matrix and the collapse of the algorithm. This requirement may not be possible to satisfy in a real remote sensing scenario, such as in the case of unexplored regions of Earth, or in planetary missions. Likewise, favorite clustering approaches such as pairwise ratios, scatttergams, or Principle Component (PCA) projections become impractical for hundreds of dimensions. Dimensionality reduction (or feature extraction) prior to classification is often performed in order to scale down the data to traditional methods. However, it is not clear how feature extraction can be done in order to preserve relevant information (class distinctions) captured by modern imaging spectrometers. The intrinsic (spectral) dimensionality of these images is a subject of open research. Dimensionality reduction is most frequently attempted by PCA or wavelets, or by selection of important image bands by domain experts. We found undesirable loss of class distinction with all of these methods [5], [6], [7]. Non-linear dimensionality reductions such as by [8] may retain more of the relevant information but systematic studies do not exist to show their general power for many classes with slight spectral shape differences, and that the discovery potential of small "interesting" groups of data is preserved. See [7] for a more detailed review.

In any case, finding optimal feature extraction for every data set requires great preprocessing efforts. A better alternative is to develop capabilities to handle the full spectral dimension. This is important for discovery as well as for benchmarking: once we know the level of useful detail that can be extracted from a high-dimensional data set using the full dimensionality, feature extraction and dimensionality reduction algorithms can be developed and tested against that knowledge. Another, fundamental aspect is that linear approaches such as PCA and/or methods based on only second order (Gaussian) statistics may not detect some of the most interesting features because hyperspectral images are characterized by higher order statistics [9]. More powerful, non-linear methods

are needed for adequate knowledge extraction.

II. COMPUTATIONAL INTELLIGENCE SOLUTIONS

A. Artificial Neural Networks

Artificial Neural Networks (ANNs) are massively parallel, finely distributed learning machines that learn to solve problems from examples. The interest in using ANNs is a) for their extreme speed when implemented in massively parallel hardware; b) for their smart brain-like data processing. This paper deals with the second aspect. A generic supervised ANN architecture is in Fig. 2, top left. It consists of layers of many simple Processing Elements (PEs) and the layers are fully connected. The connection strength, called weight, between any two PEs changes during an iterational learning process:

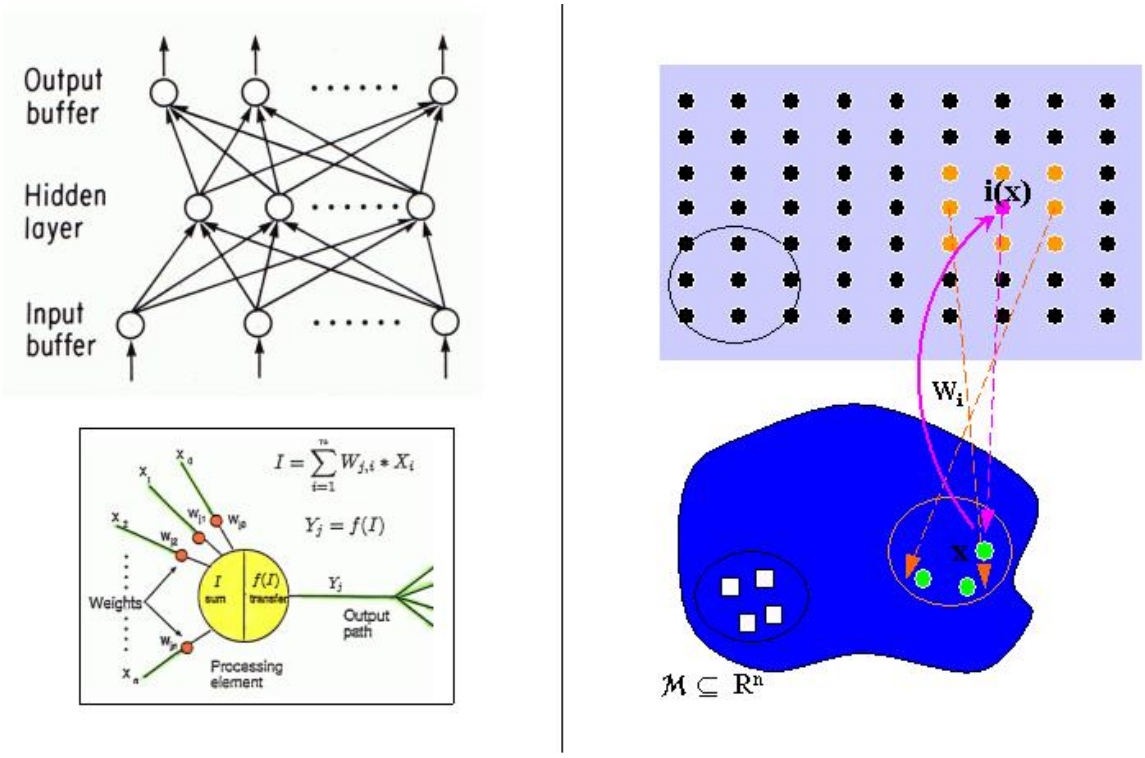

Fig. 2. **Top Left:** A generic supervised ANN scheme. Small cricles indicate Processing Elements (PEs, or artificial neurons), arrows represent the weighted connections between PEs. **Bottom Left:** The structure of a single PE. Each PE performs a very simple fundction. **Right:** Self-Organizing Map scheme. The top rectangle represents PEs arranged in a regular 2-dimensional SOM lattice. Each PE has a weight vector of the same dimension as the input space, which points to the input space. The weight vector of a PE is a prototype of all input data points that are closest to that weight vector. See description of how an SOM works, in the text.

Data samples are shown many times at the input layer, passed through the subsequent layers in which all neurons perform a simple weighted summation of their inputs. The weights are adapted at each iteration (after showing one input sample) according to a *learning rule* and based on feedback from a teacher, so that the network gradually approximates the correct answers at the output layer. (Samples from pattern classes are input as n-dimensional vectors, and class labels are encoded as m-dimensional vectors at the output layer.) ANNs can learn input-output mappings that are too complicated to describe analytically, or that have no analytical description. A comprehensive textbook on ANNs is, *e.g.,* [10]. ANNs have been used by many authors to classify spectral imagery (see [7] for some references). While most applications successfully accomplish a more accurate classification than traditional counterparts, the majority of these studies deal with relatively low dimensional data (*e.g.,* Landsat TM, Fig. 1, right panel). The most frequently used supervised ANN classifier, Back Propagation (BP), is very powerful and

fairly well understood. However, a BP that is complex enough to classify a hyperspectral image into many classes (*i.e.,* hunderds of input neurons, several dozens output neurons and many hidden neurons, in possibly multiple hidden layers) is hard to train. In what follows, we discuss Self-Organizing neural networks, which can handle high-dimensional data, and can produce a faithful representation of the structure of a complex data space to allow detailed clustering and discovery of interesting relationships in the data. A converged SOM can also be used as the hidden layer of a supervised BP-like ANN, in which the SOM greatly helps fast and easy convergence, and by virtue of its own preformed view of the data structure, prevents learning of inconsistent labels (teacher errors) thus helping highly accurate classification. Analyses of hyperspectral data from Mars and asteroids using such hybrid ANNs, and details of the hybrid architecture, are presented by [5], [11], [12]. Each resulted in improved scientific information extraction compared to previous analyses with conventional methods.

B. The Self-Organizing Map neural paradigm

Self-Organizing Maps (SOMs), invented by Teuvo Kohonen [13], are the most widely used neural maps for topographic vector quantization (faithful modeling of the structure of an input data space). They are intended to mimic the topological information mapping by biological neural maps that have been observed to form in various areas of the cerebral cortex [13]. Neural maps map data vectors $\mathbf{x}$ sampled from a data manifold $\mathcal{M} \subseteq \Re^{D_\mathcal{M}}$ of dimension $D_\mathcal{M}$, onto a discrete set $\mathcal{A}$ of neurons or processing elements (PEs), denoted by (generally multi-dimensional) indices $\mathbf{r}$. Each PE has a weight vector $\mathbf{w_r} \in \Re^{D_\mathcal{M}}$. The vector quantization is realized by a mapping $\Phi_{\mathcal{M}\to\mathcal{A}}$ through a winner-take-all rule: an input vector $\mathbf{x} \in \mathcal{M}$ is mapped onto neuron $\mathbf{i} \in \mathcal{A}$ whose pointer $\mathbf{w_i}$ is closest to $\mathbf{x}$:

$$\Phi_{\mathcal{M}\to\mathcal{A}} : \mathbf{x} \to \mathbf{i}(\mathbf{x}) = \underset{\mathbf{r}\in\mathcal{A}}{argmin} \quad d(\mathbf{x}, \mathbf{w_r}), \qquad (1)$$

where $d(\mathbf{x}, \mathbf{w_r})$ is a distance, most often the Euclidean. The neuron $\mathbf{i}$ is called the *winner*. A reverse mapping is defined as $\Phi_{\mathcal{A}\to\mathcal{M}} : \mathbf{r} \to \mathbf{w_r}$. The two mappings together determine the neural map

$$\Psi = (\Phi_{\mathcal{M}\to\mathcal{A}}, \Phi_{\mathcal{A}\to\mathcal{M}}) \qquad (2)$$

realized by the network (Fig. 2, right panel).

In the *Self-Organizing Map* [13] the neurons are arranged on a prefixed grid $\mathcal{A}$, usually a $D_\mathcal{A}$–dimensional hypercube, whose vertices are $\mathbf{r} = (r_1, \ldots, r_{D_\mathcal{A}})$. $D_\mathcal{A} = 2$ is the most popular SOM grid dimension, for its advatageous visualization. Other ordered grids are also used [13].

In neural maps the pointers are adapted during many steps of iterational learning to achieve an optimal set of prototype vectors for the quantization of the input data, that is, to match the input *pdf* as closely as possible. At each time step, a data point $\mathbf{x} \in \mathcal{M}$ randomly selected from the input distribution $\mathcal{P}(\mathbf{x})$ is presented to the map, the winning neuron $\mathbf{i}$ is determined according to (1), and the weight $\mathbf{w_i}$ along with the weights of neurons in the winner's *neighborhood* are adapted, shifted toward $\mathbf{x}$:

$$\Delta\mathbf{w_r} = \epsilon h_\lambda(\mathbf{r}, \mathbf{i}(\mathbf{x}))(\mathbf{x} - \mathbf{w_r}) \qquad (3)$$

where the *neighborhood function* $h_\lambda(\mathbf{r}, \mathbf{i}(\mathbf{x}))$ determines the extent of the adaptation for weights $\mathbf{w_r}$. ϵ is a scalar *learning rate*. The neighborhood function and the learning rate should decrease with time. $h_\lambda(\mathbf{r}, \mathbf{i}(\mathbf{x}))$ is often of Gaussian shape

$$h_\lambda(\mathbf{r}, \mathbf{i}(\mathbf{x})) = \exp\left(-\frac{(d_\mathcal{A}(\mathbf{r}, \mathbf{i}(\mathbf{x})))^2}{\lambda}\right) \qquad (4)$$

where $d_\mathcal{A}(\mathbf{r}, \mathbf{i}(\mathbf{x}))$ is a distance measure on the set $\mathcal{A}$. In the SOM usually $\lambda = 2\sigma^2$. Other neighborhood shapes, for example rectangular as in Fig. 2, right, are also used. Note that $h_\lambda(\mathbf{r}, \mathbf{i}(\mathbf{x}))$ implicitly depends on the whole set $\mathbf{W}$ of weight vectors through the winner determination of $\mathbf{i}$ according to (1). The learning described by equations (1) and (3) leads to the formation of contiguous areas of neurons in the SOM grid, whose weights collectively represent (become prototypes for) groups of similar data points of the input manifold $\mathcal{M}$. Thus clusters in $\mathcal{M}$ can be found by evaluating the SOM weights as shown in Fig. 3.

The particular definitions of the distance measures and winner selection define other neural map models such as the Neural Gas [14] that are not discussed here. Convergence of the Kohonen SOM learning has not been proven beyond 1 dimension [15] due to the fact that an energy function could not be found for the SOM [16]. For some other neural map versions this is possible (*e.g.,* [17]) but a review is beyond the scope of this paper. However, in practice convergence is usually not problematic. For more detailed overview of SOMs see [18] and [19].

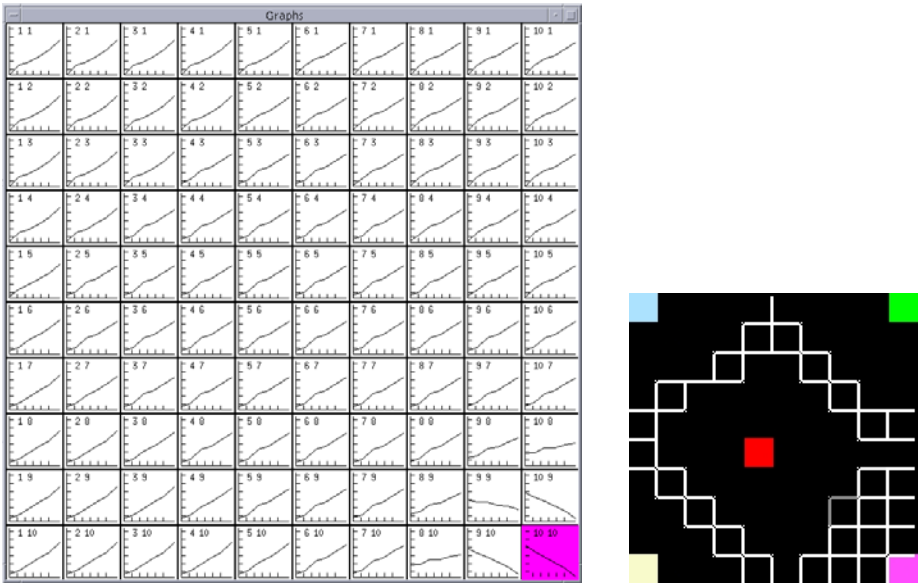

Fig. 3. Clustering a 5-class data set with SOM. **Left:** Weight vectors are shown at each neuron's location in the 10×10 SOM grid. It is easy to see that the weights adapted to the five different patterns contained in the input data set: the weight patterns form distinct groups in the four corners and in the middle of the SOM. The input data are shown in detail in [20]. Four of the five classes contain 4096 or 4095 data points while the fifth class has only one data point. This rare class, which is also dramatically different from the other four classes, is represented by only 1 weight in the lower right corner (highlighted in pink), while the large classes have more prototype weights. **Right:** Clusters identified in the SOM by visualizing the distances of the weights of adjacent neurons as fence heights, after [21]. The darker the fence between two neurons, the smaller the difference between the corresponding weights, which means that the receptive fields of those weights contain similar data. Areas of similar weights in the SOM represent clusters in the input data space. Original figure is in color.

The subset of the input space

$$V_{\mathbf{r}} = \{\mathbf{x} \in \mathcal{M} : \mathbf{r} = \Phi_{\mathcal{M}\to\mathcal{A}}(\mathbf{x})\} \tag{5}$$

mapped to neuron $\mathbf{r}$ according to (1) forms the *receptive field* of neuron $\mathbf{r}$. If the intersection of two receptive fields $V_{\mathbf{r}}$, $V_{\mathbf{r}'}$ is non-vanishing then $V_{\mathbf{r}}$ and $V_{\mathbf{r}'}$ are *neighbored*. The neighborhood relations of receptive fields form a corresponding graph structure $\mathcal{G}_{\mathcal{M}}$ in $\mathcal{A}$: two neurons $\mathbf{r}$ and $\mathbf{r}'$ are *connected* in $\mathcal{G}_{\mathcal{M}}$ if and only if their receptive fields are neighbored. The graph $\mathcal{G}_{\mathcal{M}}$ is called the *induced Delaunay-graph* ([22]). Due to the association between neurons and weight vectors $\mathcal{G}_{\mathcal{M}}$ also represents the Delaunay graph of the weights [22]. This graph can be used to evaluate the topology preserving condition of the map.

C. What is a correct SOM?

There are several interesting and very important issues related to precise approximation of an input density distribution with SOMs. First, it is important to have a topology preserving mapping Ψ, which means that the neighborhood relations between input space $\mathcal{M}$ and output space $\mathcal{A}$ should be preserved in both directions ($\Phi_{\mathcal{M}\to\mathcal{A}}$ and $\Phi_{\mathcal{A}\to\mathcal{M}}$). A mapping Ψ that is not topology preserving (that maps a group of similar input patterns to disconnected areas in the SOM grid, thus producing a "twisted map") can lead to false interpretation of data clusters from the SOM. While to some extent — especially for small input data sets — one can catch topology violations by manual evaluation of the detected clusters from a converged SOM, this is very tedious and expensive for large data sets. It is desirable to have formal tools that can alert for topology violations. The Topographic Function by Villmann *et al.*, and the Topographic Product by Bauer and Pawelzik [23], both described in [24], are such algorithms. The Growing SOM [25] provides remediation during learning.

D. Magnification in SOMs

Map magnification is another important theoretical aspect that we make use of. It refers to the relative size of SOM areas representing subsets of the input space, and is formally described by the following power law between the density of weights in the input space $Q(\mathbf{w})$ and the *pdf* $P(\mathbf{x})$ of the input samples:

$$Q(\mathbf{w}) = CP(\mathbf{x})^{\alpha} \tag{6}$$

where α is the *magnification exponent* and C is a constant [26]. As described by [26] a converged SOM with $\alpha = 1$ maximizes information theoretic entropy. $\alpha = D/(D+2)$ for D-dimensional data corresponds to minimum mean squared error quantization. $\alpha < 0$ enlarges response areas in the SOM for low-frequency inputs, which can help the discovery of low-probability events by enhancing the detectability of rare classes. Kohonen's SOM algorithm (KSOM) [13] has been shown to achieve $\alpha = 2/3$ for 1- and 2-D maps under certain conditions [27], which is optimal in neither minimum distortion nor maximum entropy sense. A heuristic SOM variant called Conscience algorithm [28] produces $\alpha = 1$ but not any other value. Bauer, Der and Herrmann proposed a principled SOM algorithm (referenced here as BDH) for the *explicit* control of the magnification exponent by using adaptively adjusted local learning rates [26]. This possibility is very attractive because it could be a tool for obtaining various quantization properties, appropriate for specific data mining purpose. However, analytical proof for successful magnification control only exists for 1-dimensional data and for 2-dimensional data whose components are statistically independent, while all interesting real data are high-dimensional with all kinds of correlations between data dimensions. Jain and Merényi presented numerical simulations in [20] on 2-dimensional simple "forbidden" data with known *pdf*, for which the achieved magnification α could be computed from the SOM. Similarly, 6-dimensional synthetic spectral images with known class structure were examined for magnification effects. These experiments indicate that the BDH algorithm may work for data unsupported by the theory. Following we show samples from analyses conducted on real spectral images with BDH, and with Conscience algorithm. We are not aware that the validity and power of the intriguing BDH scheme has been gauged for real complex data outside of our studies.

III. SOM STUDIES ON REAL SPECTRAL IMAGES

A. Discovery with negative SOM magnification

A 512 x 512 pixel, 8-band remote sensing image of Ocean City, Maryland, was used to study the effect of forced negative magnification. This data set is > 2 dimensional with high pairwise (0.5 to 0.95) correlations. The actual α value achieved by the map cannot be computed because the data distribution is unknown (see [30], [20] for computing α for known densities), but we can compare the areal representation of known small classes in the BDH SOM and in an SOM that learned with the Conscience algorithm to see if the rare classes occupy larger areas in the BDH SOM. (In [29] we verified that the Conscience algorithm does achieve $\alpha = 1$ on similar synthetic data.) We also look for previously unidentified clusters. Fig. 4 demonstrates the discovery of a small cluster. It also shows another, pale aqua class (V) that was known at the time of an earlier supervised classification, but became more pronounced by BDH clustering. Fig. 5 compares the two SOMs. Shown on the left is the 40 x 40 SOM formed by BDH learning with intended $\alpha = -0.8$, using only the upper right quadrant of the image, *i.e.,* 1/4 of the data. The newly discovered rare cluster (greenish-yellow) is indicated by the middle arrow. The spectral signature of this cluster is distinct from all others as seen in Fig. 4. Also indicated are two other small clusters that correspond to the previously known V (pale aqua) and C (white) classes from the supervised class map (Fig. 4). The 40 x 40 Conscience SOM that was learned using the *entire* image, is in the middle. The new greenish-yellow cluster was hard to see in this map, and was only found because we looked for it based on the BDH discovery. This rare cluster covers only 3 PEs in the Conscience SOM in contrast to 7 PEs in the BDH SOM where it is also more contoured by stronger "fences". Similarly, the previously knwon small V class is represented by 4 PEs in the Conscience SOM vs 6 PEs in the BDH SOM. The white class (C) occupies 4 PEs in both SOMs, in spite that within the 1/4 subimage used for BDH clustering the white class only occurs in a small rectangle (not circled) at the upper right corner, while there are many more white class pixels in the entire image used for the Conscience SOM training. These observations clearly indicate that, relative to Conscience SOM the BDH performed negative magnification.

B. Rare clusters in a Mars spectral iamge

One of the most important things in the exploration of other planets is the discovery of new, surprising or rare materials. We analyzed a spectral image of the Martian surface taken by the Imager for Mars Pathfinder (IMP) in 1997. This image is one of the so-called octants of the SuperPan (360 degree panorama) image obtained by the left eye of the IMP, consisting of 8 bands taken at wavelengths from 0.44 to 1.001 μm. The spatial size is nearly 1,000 x 1,000 pixels, with a large area occupied by the landers ramp. The Martian surface shows in about 600,000 pixels.

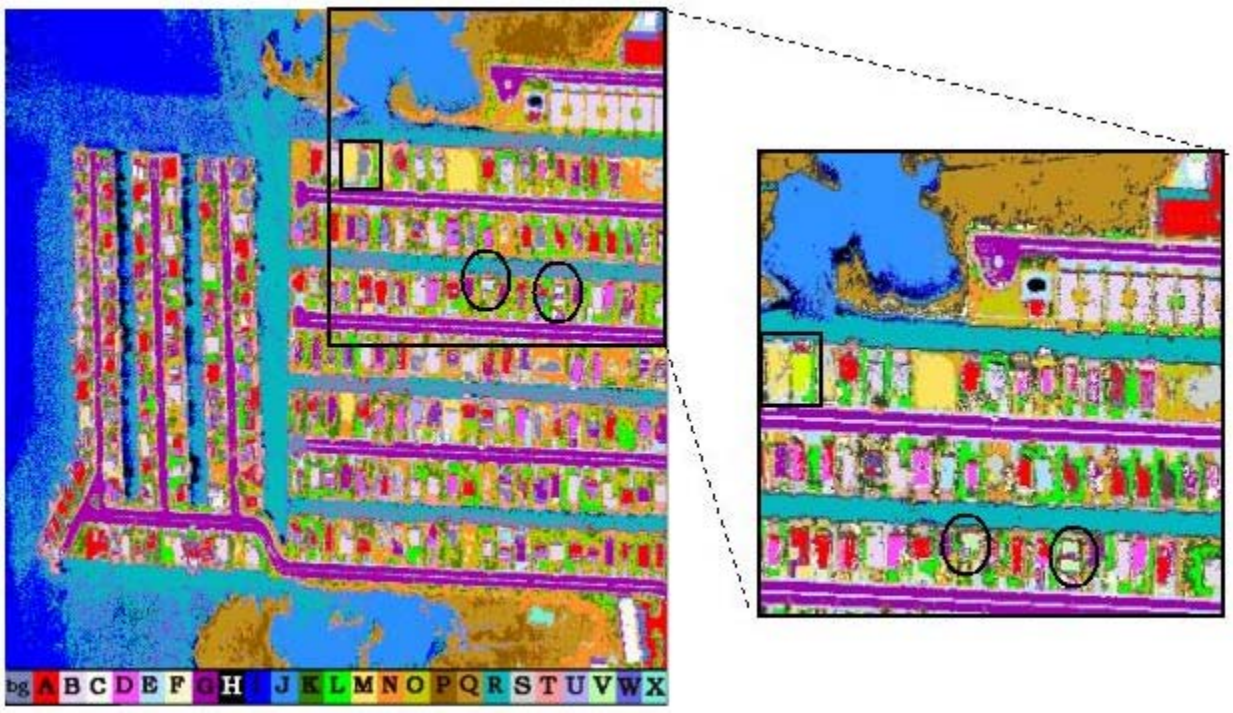

Fig. 4. Comparison of supervised classification and BDH clustering with $\alpha < 0$. **Left:** An earlier supervised classification that accurately mapped 24 known urban cover types. Centered in the small black rectangle within the framed upper right quadrant is an unclassified grey spot (the color of the background, 'bg') apparently of the shape of a building. **Right:** SOM clustering using BDH magnification control with $\alpha = -0.8$ on the upper right quadrant of the image. First, notice that the agreement between the supervised class map and this cluster map is striking, which inspires confidence in the clustering. Secondly, notice that the spot that remained unclassified in the supervised map is now filled exactly and with a new color (greenish-yellow). The spectral signature of this new cluster is distinct from all others, moreover, it only occurs at this location: we discovered a rare class! Fig. 5 shows the SOM view of this discovery, and the spectral signatures of clusters. Original figure is in color.

We clustered this image in a previous work [31] with a Conscience SOM, and found known, very rare occurrences of a black rock type that is a fairly pristine, olivine and/or pyroxene rich rock and of great interest to geologists. We also found two subtypes of the black rock, the spectral signatures of which show mineralogic distiction: one subtype has an absorption at approximately 0.93 μm (consistent with orthopyroxene), the other has an absorption at around 1 μm (consistent with clinopyroxene or olivine). One subset of the clustered image is in Fig. 6, with the inset enlarging an occurrence of both rare subclasses. The full image with class descriptions and corresponding mean spectral shapes are shown in [31].

Fig. 7 compares the Conscience SOM and by a BDH SOM of the same data. The representation areas of both clusters increased, and their separation became stronger, in the BDH SOM. Moreover, further subtypes were discovered by the BDH, within the original pink cluster (white) and in-between pink and aqua. The corresponding mean spectral signatures shown in [32] confirm these subclusters.

C. A 194-dimensional case

So far we illustrated the power of Self-Organizing Maps on relatively low-dimensional (8-band) spectral signatures. The reason is that we wanted to present "hot new" results with controlled map magnification, and our experiments with forced map magnification still need a lot of careful simulations in order to gauge the behavior of the BDH algorithm on higher-dimensional data. However, we have excellent results with Conscience SOM for 100–200 dimensional spectral data. For example, using a Conscience SOM to cluster 60-dimensional asteroid spectra helped formal identification of suspected asteroid types that could not be detected with other methods such as PCA and minimum spanning tree clustering [5]. Discovery of a

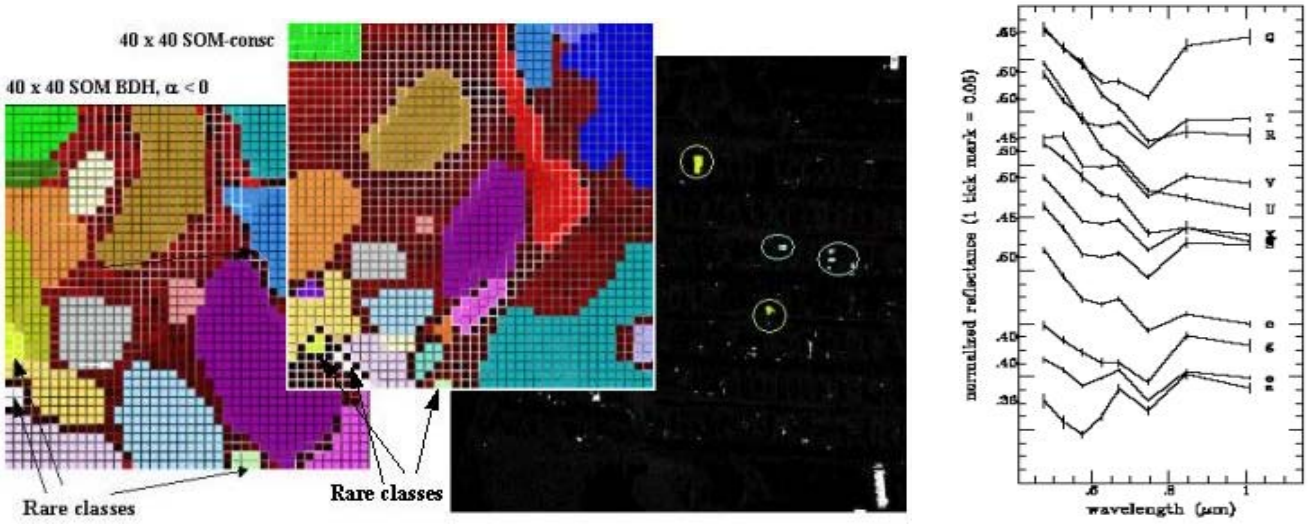

Fig. 5. **Color Panel:** Comparison of SOMs developed by BDH versus Conscience learning. *Left:* The SOM learned by BDH, $\alpha < 0$, using the upper right quadrant of the Ocean City image. *Middle:* The SOM learned by Conscience algorithm ($\approx \alpha = 1$), using the entire Ocean City image. *Right:* The rare classes in the image. It is apparent, as explained in the text, that the rare clusters are magnified in the BDH SOM in comparison to the Conscience SOM. Original figure is in color. **Graphs:** The mean spectral signatures of half of the clusters identified with BDH SOM. The spectrum of the new greenish-yellow cluster is the bottom one, and it is very different from all others (including the other half of the clusters not shown because of space considerations).

new Martian soil type was made from 100-dimensional data through the use of an hybrid ANN with a Conscience SOM as the hidden layer [33]. Perhaps the most challenging data sets we dealt with are AVIRIS images. [7] presents segmentation of a 194-band image into 32 clusters based on the 194-dimensional spectral signatures. For lack of space the reader is referred to that paper for the figures. Besides showing high degree of agreement with the geologist's knowledge, this cluster map also produced discoveries, surface units that the geologist was unaware of! [7] also demontstrates, on the same image, a highly accurate supervised classification using the above mentioned hybrid ANN, and how the high data dimension leads to poor performance by a Maximum Likelihood classifier. Since we could verify the clusters (the data distribution) for this image based on ground truth, we could make a crude estimation of the magnification exponent to see if the Conscience algorithm produces, as advertised, a linear relationship ($\alpha = 1$) between weight and data density for such high dimensional data. Fig. 6 of [7] shows that to be the case.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

This paper argued that for complex scientific data such as remote sensing spectral images, precise clustering and classification pose unique challenges that call for computational intelligence tools. Artificial Neural Networks and in particular, Self-Organizing Maps and advanced variants were described, and their capabilities in information extraction from spectral images were demonstrated through analyses of real Earth and space science data. A new line of study, assessment of the scope of the explicit magnification control algortihm by [26] was presented along with successful applications. This numerical validation of the BDH algorithm is ongoing, and it needs much more work to understand its use and applicability to higher dimensional data. It is hoped that these simulations will lead to increasingly more sophisticated and more powerful analyses of intricate, large-volume data. We want to point out that microspectral imaging (spectral imaging through a microscope) has been gaining recognition in the past few years for distinguishing pathologies in biological tissues based on spectral properties. This opens up a large field of extremely important medical applications such as identification of cancer cells on the basis of their chemical composition rather than from morphological features (which is a much more subjective judgement), and thus has the potential of automating parts of the screening process. The techniques we elaborated on are directly applicable to such medical imagery.

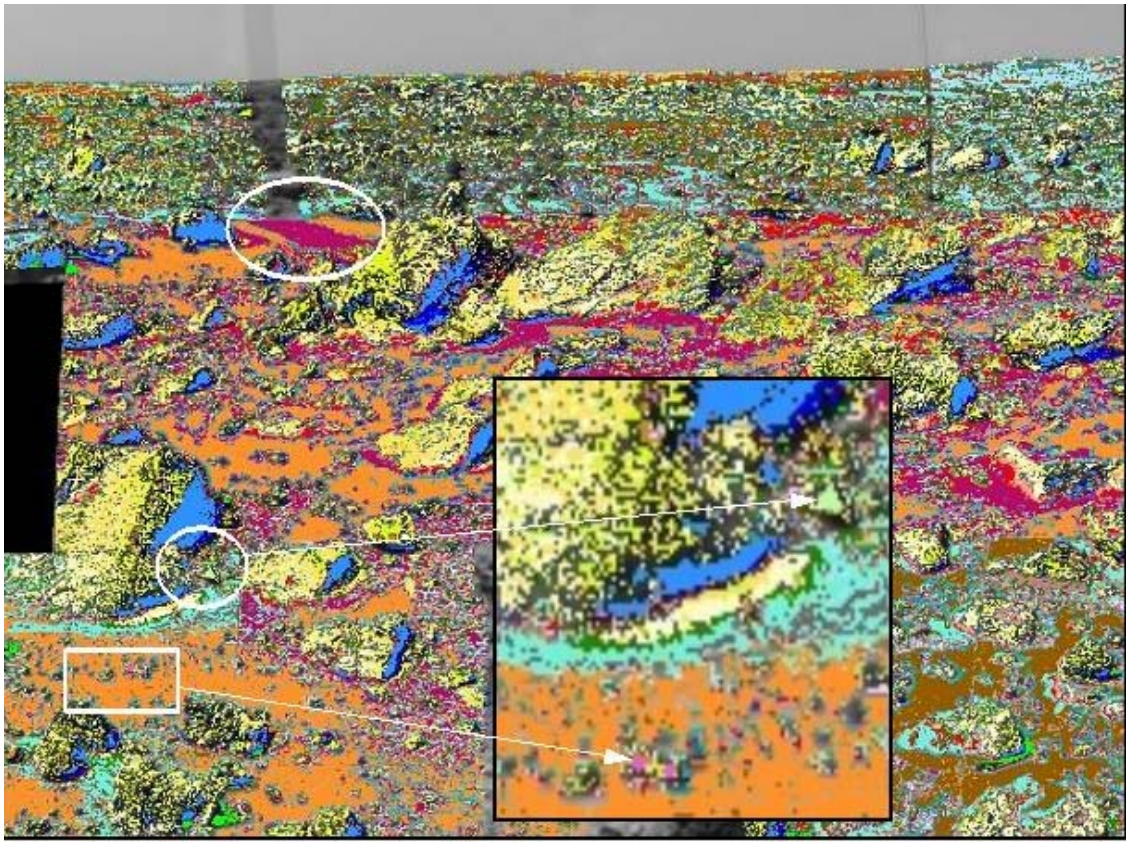

Fig. 6. Clusters obtained with a Conscience SOM, from IMP SuperPan octant S0184 of the Martian surface at the Mars Pathfinder landing site. The full image and discussion of clusters is given in [31]. Here we focus on two extremely rare clusters that are subtypes of a geologically relevant "black rock" material: the pink class and the pale aqua class within the white rectangle and the white small oval, respectively. The same are enlarged and pointed at by arrows in the inset. These occurrences contain less than 25 pixels each.

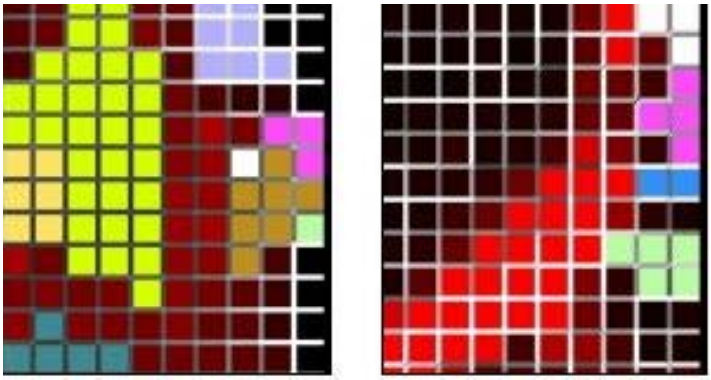

Fig. 7. **Left:** Detail from the 40 x 40 Conscience SOM used in [31], showing the representation of the pink and aqua "black rock" clusters from Fig. 6. **Right:** Detail from the 40 x 40 BDH SOM showing the representation of the same two rare clusters. The separation between pink and aqua clusters is much more definite than in the Conscience SOM. The areal representation of the aqua class is much larger in the BDH SOM. The original pink type has further split into subtypes and is now represented by the pink and white areas together, doubling the representation area (The white color was recycled here and is not the same spectral type as in the Conscience SOM).

Acknowledgments

This research was partially supported by grant NAG9-10432 from the Applied Information Systems Research Program, and by grant NAG5-13294 from the Mars Data Analysis Program, both under NASA's Office of Space Science. Contributions by the following colleagues are gratefully acknowledged: Dr. Thomas Villmann, University of Leipzig, Germany;

Dr. William H. Farrand, Space Science Institute, Boulder, CO, USA; Ms. Abha Jain, former M.Sc. student at Rice University; Dr. Bea Csathó, Byrd Polar Research Institute at the Ohio State University. Thanks to Mr. Philip Tracadas, Rice University, for software development, and to Dr. Scott Murchie, Johns Hopkins Applied Physics Lab, for the latest calibrated version of the Mars SuperPan imagery.

References

[1] R. O. Green, "Summaries of the 6th Annual JPL Airborne Geoscience Workshop, 1. AVIRIS Workshop," Pasadena, CA, March 4–6 1996.

[2] R. W. Basedow, D. C. Carmer, and M. L. Anderson, "HYDICE: An airborne system for hyperspectral imaging," in *Proc. SPIE*, Orlando, FL, 17-18 April 1995, vol. 2480, pp. 258–267.

[3] J. Campbell, *Introduction to Remote Sensing*, The Guilford Press, U.S.A., 1996.

[4] R. N. Clark, "Spectroscopy of rocks and minerals, and principles of spectroscopy," in *Manual of Remote Sensing*, A. Rencz, Ed. Wiley and Sons, New York, 1999.

[5] E. S. Howell, E. Merényi, and L. A. Lebofsky, "Classification of asteroid spectra using a neural network," *Jour. Geophys. Res.*, vol. 99, no. E5, pp. 10,847–10,865, 1994.

[6] T. Moon and E. Merényi, "Classification of hyperspectral images using wavelet transforms and neural networks," in *Proc. Annual SPIE Conf.*, San diego, CA, 1995, p. 2569.

[7] E. Merényi, "Precision mining of high-dimensional patterns with self-organizing maps: Interpretation of hyperspectral images.," in *Quo Vadis Computational Intelligence: New Trends and Approaches in Computational Intelligence (Studies in Fuzziness and Soft Computing, Vol 54, P. Sincak and J. Vascak Eds.)*. 2000, Physica Verlag.

[8] J. A. Benediktsson J. R. Sveinsson and et al., "Classification of very-high-dimensional data with geological applications," in *Proc. MAC Europe 91*, Lenggries, Germany, 1994, pp. 13–18.

[9] E. Merényi and T. Villmann, "Self-Organizing Neural Net Approaches For Hyperspectral Images," in *Proc. Int'l Conf. on Intelligent Computing and Infromation Systems*, Ain Shams University, Cairo, Egypt, June 24–26 2002.

[10] S. Haykin, *Neural Networks. A Comprehensive Foundation*, McMillan, New Jersey, 1999.

[11] E. Merényi, E. S. Howell, and et al., "Prediction of water in asteroids from spectral data shortward of 3 microns," *ICARUS*, vol. 129, pp. 421–439, 1997.

[12] E. Merényi, R. B. Singer, and J. S. Miller, "Mapping of spectral variations on the surface of Mars from high spectral resolution telescopic images," *ICARUS*, vol. 124, pp. 280–295, 1996.

[13] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin Heidelberg New York, 1997.

[14] Th. Martinetz, S.G. Berkovich, and K. Schulten, "Neural gas network for vector quantization and its application to time-series prediction," *IEEE Trans. on Neural Networks*, vol. 4(4), pp. 558–569, 1993.

[15] M. Cottrell, J.C. Fort, and G. Pages, "Theoretical aspects of the SOM algorithm," *Neurocomputing*, vol. 21, pp. 119–138, 1998.

[16] E. Erwin, K. Obermayer, and K. Schulten, "Self-organizing maps: ordering, convergence properties and energy functions," *Biol. Cyb.*, vol. 67, pp. 47–55, 1992.

[17] T. Heskes, "Energy functions for self-organizing maps," in *Kohonen Maps*, E. Oja and S. Kaski, Eds., Amsterdam (Holland), June 1999, Helsinki, pp. 303–315, Elsevier.

[18] T. Villmann, E. Merényi, and B. Hammer, "Neural maps in remote sensing image analysis," *Neural Networks*, vol. 16, pp. 389–403, 2003.

[19] T. Villmann and E. Merényi, "Extensions and modifications of the Kohonen-SOM and applications in remote sensing image analysis," in *Self-Organizing Maps: Recent Advances and Applications (U.Seiffert and L.C. Jain Eds.)*. 2001, pp. 121–145, Springer-Verlag.

[20] A. Jain and E. Merényi, "Forbidden magnification? I.," in *Proc. Of European Symposium on Artificial Neural Networks (ESANN'04)*, Brussels, Belgium, 2004, D facto publications.

[21] A. Ultsch and H. P. Simeon, "Kohonen's self organizing feature map for exploratory data analysis," in *Proc. INNC-90-PARIS* **I**, Paris, 1990, pp. 305–308.

[22] Th. Martinetz and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7(3), pp. 507–522, 1994.

[23] H.-U. Bauer and K.R. Pawelzik, "Quantifying the neighborhood preservation of Self-Organizing Feature Maps," *IEEE Trans. on Neural Networks*, vol. 3, pp. 570–579, 1992.

[24] Th. Villmann, R. Der, M. Herrmann, and Th. Martinetz, "Topology Preservation in Self–Organizing Feature Maps: Exact Definition and Measurement," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 256–266, 1997.

[25] H.-U. Bauer and Th. Villmann, "Growing a Hypercubical Output Space in a Self–Organizing Feature Map," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 218–226, 1997.

[26] H.-U. Bauer, R. Der, and M. Herrmann, "Controlling the magnification factor of self–organizing feature maps," *Neural Computation*, vol. 8, no. 4, pp. 757–771, 1996.

[27] H. Ritter and K. Schulten, "On the stationary state of Kohonen's self-organizing sensory mapping," *Biol. Cyb.*, vol. 54, pp. 99–106, 1986.

[28] D. DeSieno, "Adding a conscience to competitive learning," in *Proc. ICNN, July 1988* **I**, New York, 1988, pp. 117–124.

[29] E. Merényi and A. Jain, "Forbidden magnification? II.," in *Proc. Of European Symposium on Artificial Neural Networks (ESANN'04)*, Brussels, Belgium, 2004, D facto publications.

[30] H. Ritter, "Asymptotic level density for a class of vector quantization processes," *IEEE Trans. on Neural Networks*, vol. 2, pp. 173–175, 1991.

[31] E. Merényi, W.H. Farrand, and P. Tracadas, "Mapping surface materials on Mars from Mars Pathfinder spectral images with HYPEREYE.," in *Proc. International Conference on Information Technology (ITCC 2004)*, Las Vegas, Nevada, 2004, pp. 607–614, IEEE.

[32] E. Merényi, A. Jain, and W.H. Farrand, "Applications of SOM magnification to data mining," *WSEAS Trans. on Systems*, vol. 3, no. 5, pp. 2122–2128, 2004.

[33] E. Merényi, K. S. Edgett, and R. B. Singer, "Deucalionis Regio, Mars: Evidence for a new type of immobile weathered soil unit," *ICARUS*, vol. 124, pp. 296–307, 1996.