

Extensions and Modifications of the Kohonen-SOM and Applications in Remote Sensing Image Analysis*

Thomas Villmann¹ and Erzsébet Merényi²

¹ University Leipzig,
Clinic of Psychotherapy
Karl-Tauchnitz-Str. 25
04107 Leipzig, Germany
Email: villmann@informatik.uni-leipzig.de

² Rice University
Department of Electrical and Computer Engineering
6100 Main Street, MS 380
Houston, Texas, U.S.A.
Email: erzsebet@ece.rice.edu

Summary. Utilization of remote sensing multi- and hyperspectral imagery has shown a rapid increase in many areas of economic and scientific significance over the past ten years. Hyperspectral sensors, in particular, are capable of capturing the detailed spectral signatures that uniquely characterize a great number of diverse surface materials. Interpretation of these very high-dimensional signatures, however, has proved an insurmountable challenge for many traditional classification, clustering and visualization methods. This chapter presents spectral image analyses with Self-Organizing Maps (SOMs). Several recent extensions to the original Kohonen SOM are discussed, emphasizing the necessity of faithful topological mapping for correct interpretation. The effectiveness of the presented approaches is demonstrated through case studies on real-life multi- and hyperspectral images.

1 Introduction

Airborne and satellite-borne remote sensing spectral imaging has become one of the most advanced tools for collecting vital information about the surface 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 (TM)*). 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

* in: U.Seiffert and L.C. Jain (Eds.), *Self-Organizing Maps. Recent Advances and Applications*. Springer-Verlag Berlin, p. 121–145, 2001.

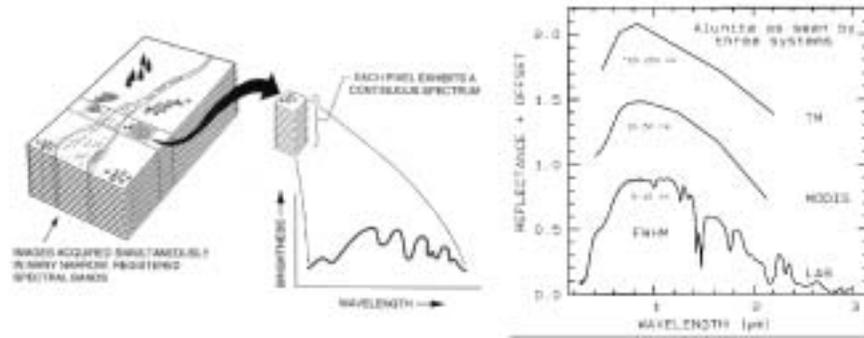


Fig. 1. Left: The concept of hyperspectral imaging. Figure from [9]. Right: The spectral signature of the mineral alunite as seen through the 6 broad bands of Landsat TM, as seen by the moderate spectral resolution sensor MODIS (20 bands in this region), and as measured in laboratory. Figure from [10]. Hyperspectral sensors such as AVIRIS of NASA/JPL [16] produce spectral details comparable to laboratory measurements.

0.4 – 2.5 μm range in the case of Visible and Near-Infrared imaging) with several broad bandpasses, leaving large gaps between the bands, imaging spectrometers sample a spectral window contiguously with very narrow, 10 – 20 nm bandpasses. Hyperspectral technology is in great demand because direct identification of surface compounds is possible without prior field work, for materials with known spectral signatures.

Spectral images consist of an array of multi-dimensional vectors assigned to particular spatial areas (pixel locations) reflecting the response of a spectral sensor at various wavelengths (see Fig. 1). These vectors are called spectra. A spectrum is a characteristic pattern that provides a clue to the surface material within the respective area. Depending on the wavelength resolution and the width of the wavelength window used by a particular sensor, the dimensionality of the spectra can be as low as 5 – 6 (such as in Landsat TM), or as high as several hundred for hyperspectral imagers.

Classification of intricate, high-dimensional spectral signatures has turned out far from trivial. Discrimination among many surface cover classes, discovery of spatially small, interesting spectral species proved to be an insurmountable challenge to many traditional clustering and classification methods. This motivates research into advanced and novel approaches. [30]. By customary measures (such as, for example, Principle Component Analysis (PCA)) the intrinsic spectral dimensionality of hyperspectral images appears to be surprisingly low, 5 – 10 at most. Yet dimensionality reduction to such low numbers has not been successful in terms of preservation of important class distinctions. The spectral bands, many of which are highly correlated, may lie on a low-dimensional but non-linear manifold, which is a scenario that eludes many classical approaches.

One powerful approach to these problems is the application of Self-Organizing Maps (SOMs) [24] to implement a suitable mapping procedure which should end in a topology preserving projection of the high-dimensional data onto a low-dimensional lattice. In most applications a two-dimensional SOM lattice is the common choice of lattice structure because of its easy visualization [23]. In general, this choice is not guaranteed to produce a topology preserving mapping and the interpretation of the resulting map may fail [40]. Topology preserving mapping, informally speaking, means that similar data vectors are mapped onto the same or neighbored locations in the lattice and vice versa [41]. A further aspect that should be addressed here is the problem of the detection of rarely occurring surface material classes in the images. If the SOM is used as a classifier system the distribution of the weight vectors across the SOM lattice is determined by a power function of the probability density of the data vectors [37], with the so-called magnification factor as the power. As we will see later, for magnification factors less than 1, which is the case for the standard SOM, it may be difficult or impossible to separate spectral patterns of seldom occurring surface materials.

In the last few years extensions of the standard SOM were derived to respond to the above challenges. As shown in [3] it is possible to control the magnification of the SOM by the introduction of a local learning rate that is dependent on the data density. The corresponding learning scheme can easily be implemented into the standard learning rule. A growing SOM (GSOM) approach was developed to generate a guaranteed topology preserving mapping in a simple hypercube structure of the lattice [5]. Both the GSOM and magnification control approaches were shown to be powerful instruments for visualization and classification of remote sensing spectral data.

This chapter is organized as follows: Sec. 2 is a short overview of the above extensions of the standard SOM. In section 3 the data sets are described. Finally, in section 4 application results are presented and discussed.

2 SOMs and Data Mapping

Self-organizing maps [24] as a special kind of neural maps project data from some (possibly high-dimensional) input space $\mathcal{V} \subseteq \mathbb{R}^{D_{\mathcal{V}}}$ onto a position in some output space (neural map) \mathcal{A} , such that a continuous change of a parameter of the input data should lead to a continuous change of the position of a localized excitation in the neural map. This property of *neighborhood preservation* depends on an important feature of the SOM, its output space topology, which has to be specified prior to learning. If the topology (i.e. dimensionality and edge length ratios) of \mathcal{A} does not match that of the data shape, neighborhood violations are likely to occur [43]. This can be cast in a formal way by writing the output space positions as $\mathbf{r} = (i_1, i_2, i_3, \dots, i_{n_j})$, $1 < i_k < n_k$ with $N = n_1 \times n_2 \times \dots \times n_j$ where n_k , $k = 1, \dots, j$ is the dimension of \mathcal{A} (the length of the edge of the lattice) in the k^{th}

direction.¹ Associated with each neuron $\mathbf{r} \in \mathcal{A}$, is a weight vector, or *pointer*, \mathbf{w}_r in \mathcal{V} . The mapping $\Psi_{\mathcal{V} \rightarrow \mathcal{A}}$ is realized by a winner take all rule

$$\Psi_{\mathcal{V} \rightarrow \mathcal{A}} : \mathbf{v} \mapsto \mathbf{s} = \underset{\mathbf{r} \in \mathcal{A}}{\operatorname{argmin}} \|\mathbf{v} - \mathbf{w}_r\| \quad (1)$$

whereas the reverse mapping is defined as $\Psi_{\mathcal{A} \rightarrow \mathcal{V}} : \mathbf{r} \mapsto \mathbf{w}_r$. The two functions together determine the map

$$\mathcal{M} = (\Psi_{\mathcal{V} \rightarrow \mathcal{A}}, \Psi_{\mathcal{A} \rightarrow \mathcal{V}}) \quad (2)$$

realized by the SOM network. All data points $\mathbf{v} \in \mathbb{R}^{D_{\mathcal{V}}}$ that are mapped onto the neuron \mathbf{r} make up its receptive field $\hat{\Omega}_r$. The masked receptive field of neuron \mathbf{r} is defined as the intersection of its receptive field with \mathcal{V} :

$$\Omega_r = \{\mathbf{v} \in \mathcal{V} : \mathbf{r} = \Psi_{\mathcal{V} \rightarrow \mathcal{A}}(\mathbf{v})\}. \quad (3)$$

Therefore, the masked receptive fields Ω_r are closed sets. All masked receptive fields form the Voronoi tessellation of \mathcal{V} . If the intersection of two masked receptive fields $\Omega_r, \Omega_{r'}$ is non-vanishing we call Ω_r and $\Omega_{r'}$ neighbored. The neighborhood relations form a corresponding graph structure $\mathcal{G}_{\mathcal{V}}$ in \mathcal{A} : two neurons are connected in $\mathcal{G}_{\mathcal{V}}$ if and only if their masked receptive fields are neighbored. The graph $\mathcal{G}_{\mathcal{V}}$ is called the induced Delaunay-graph (See, for example, [27] for detailed definitions). Due to the bijective relation between neurons and weight vectors, $\mathcal{G}_{\mathcal{V}}$ also represents the Delaunay graph of the weights.

To achieve the map \mathcal{M} , SOMs adapt the pointer positions during the presentation of a sequence of data points $\mathbf{v} \in \mathcal{V}$ selected from a data distribution $\mathcal{P}(\mathcal{V})$, as follows:

$$\Delta \mathbf{w}_r = \epsilon h_{rs}(\mathbf{v} - \mathbf{w}_r). \quad (4)$$

h_{rs} is the neighborhood function, usually chosen to be of Gaussian shape:

$$h_{rs} = \exp\left(-\frac{\|\mathbf{r} - \mathbf{s}\|^2}{2\sigma^2}\right) \quad (5)$$

Note that h_{rs} is dependent on the best matching neuron (1).

Topology preservation in SOMs is defined as the preservation of the continuity of the mapping from the input space onto the output space, more precisely it is equivalent to the *continuity* of \mathcal{M} between the *topological spaces* with properly chosen metric in both \mathcal{A} and \mathcal{V} . For lack of space we refer to [43] for detailed considerations. The topology preserving property can be used for immediate evaluations of the resulting map, for instance for interpretation as a color space, as demonstrated in sec. 4.2. Topology preservation also allows the applications of interpolating schemes such as the parametrized SOM (PSOM) [36] or interpolating SOM (I-SOM) [14]. A higher degree of topology preservation, in general, improves the

¹ Other spatial arrangements are also possible, which can be described by a connectivity matrix. Here we only consider hypercubes.

accuracy of the map [4]. As pointed out in the introduction violations of topographic mapping can result in false interpretations. Several approaches were developed to judge the degree of topology preservation for a given map. Here we briefly describe a variant \tilde{P} of the well known topographic product P [4]. Instead of the Euclidean distances between the weight vectors, this measure uses the respective distances $d^{\mathcal{G}_V}(\mathbf{w}_r, \mathbf{w}_{r'})$ of minimal path lengths in the induced Delaunay-graph \mathcal{G}_V of the \mathbf{w}_r . During the computation of \tilde{P} for each node \mathbf{r} the sequences $\mathbf{n}_j^{\mathcal{A}}(\mathbf{r})$ of j -th neighbors of \mathbf{r} in \mathcal{A} and $\mathbf{n}_j^{\mathcal{V}}(\mathbf{r})$ describing the j -th neighbor of \mathbf{w}_r , have to be determined. These sequences and further averaging over neighborhood orders j and nodes \mathbf{r} finally leads to

$$\tilde{P} = \frac{1}{N(N-1)} \sum_{\mathbf{r}} \sum_{j=1}^{N-1} \frac{1}{2j} \log(\Theta) \quad (6)$$

with

$$\Theta = \prod_{i=1}^j \frac{d^{\mathcal{G}_V}(\mathbf{w}_r, \mathbf{w}_{\mathbf{n}_i^{\mathcal{A}}(\mathbf{r})})}{d^{\mathcal{G}_V}(\mathbf{w}_r, \mathbf{w}_{\mathbf{n}_i^{\mathcal{V}}(\mathbf{r})})} \cdot \frac{d_{\mathcal{A}}(\mathbf{r}, \mathbf{n}_i^{\mathcal{A}}(\mathbf{r}))}{d_{\mathcal{A}}(\mathbf{r}, \mathbf{n}_i^{\mathcal{V}}(\mathbf{r}))}. \quad (7)$$

\tilde{P} can take on positive or negative values: if $\tilde{P} < 0$ holds the output space is too low-dimensional, and for $\tilde{P} > 0$ the output space is too high-dimensional. In both cases neighborhood relations are violated. Only for $\tilde{P} \approx 0$ does the output space approximately match the topology of the input data.²

Application of SOMs to very high-dimensional data can produce difficulties which may result from the so-called '*curse of dimensionality*': the problem of sparse data caused by the large data dimensionality. We want to refer to two methods for overcoming this problem. The first approach, introduced by KASKI [23] uses the fact that in extremely high-dimensional data spaces the inner product of vectors often tends to be zero, i.e. many of the data vectors $\mathbf{v} \in \mathcal{V}$ seem to be nearly orthogonal to one another. Assume that the data undergo a (linear) random mapping

$$\bar{\mathbf{x}} = \mathbf{R}\mathbf{x} \quad (8)$$

where \mathbf{R} is a random matrix the columns of which are normalized to unity and the components of each column are independent, identically and normally distributed with zero mean. Let $d_{\mathbf{R}}$ be the *reduced* dimension of the target space. If we assume that $d_{\mathbf{R}}$ is large and consider then $\mathbf{R}^T \mathbf{R} = \mathbf{I} + \epsilon$ with the identity matrix \mathbf{I} , one can show that the elements ϵ_{ij} are approximately normally distributed and the respective variance, denoted by σ_{ϵ}^2 , can be approximated by $\sigma_{\epsilon}^2 \approx \frac{1}{d}$. Considering the inner

² The present variant \tilde{P} overcomes the problem of strongly curved maps which may be judged neighborhood violating by the original P even though the shape of the map might be perfectly justified [43].

product $\bar{\mathbf{x}}^T \bar{\mathbf{y}}$ of two mapped vectors $\bar{\mathbf{x}}, \bar{\mathbf{y}}$ we have for the deviation δ from the original inner product $\mathbf{x}^T \mathbf{y}$ the expectation value zero and the variance

$$\sigma_\delta^2 = \left[1 + \left(\sum_k x_k y_k \right)^2 - 2 \sum_k x_k^2 y_k^2 \right] \sigma_\epsilon^2 \quad (9)$$

Thus, random mapping can be applied to reduce the data dimension if the original dimension is very large. We should emphasize that the remaining dimension $d_{\mathbf{R}}$ must be large.³ Hence, this scheme is not practicable in many applications for final processing but it can serve as a useful preprocessing. In particular, in the remote sensing application presented below neither the assumptions of the zero-tendency of the inner product nor the requirement of high reduced data dimensionality is not fulfilled.

The second method introduces a local neighborhood range σ_r for sparse data sets in (5) to obtain a faithful mapping [11]. σ_r is determined by the inner curvature strength γ of the lattice which can be obtained by a wavelet analysis of the neuron weights. For a more detailed description we refer to [11].

2.1 Structure Adaptation by GSOM

The growing SOM (GSOM) approach [5] is an extension of the standard SOM. Its output is a *structure adapted* hypercube \mathcal{A} , produced by adaptation of both the dimensions and the respective edge length ratios of \mathcal{A} during the learning, in addition to the usual adaptation of the weights. In comparison to the standard SOM, the overall dimensionality and the dimensions along the individual directions in \mathcal{A} are variables that evolve into the hypercube structure most suitable for the input space topology. The GSOM starts from an initial 2-neuron chain, learns like a regular SOM, adds neurons to the output space based on the criterion described below, learns again, adds again, etc., until a prespecified maximum number N_{\max} of neurons is distributed over \mathcal{A} . The output space topology always remains of the form $n_1 \times n_2 \times \dots$, with $n_j = 1$ for $j > D_{\mathcal{A}}$, where $D_{\mathcal{A}}$ is the current dimensionality of \mathcal{A} . Hence, the initial configuration is $2 \times 1 \times 1 \times \dots$, $D_{\mathcal{A}} = 1$. From there it can grow either by adding nodes in one of the directions that are already included in the output space or by initializing a new dimension. This decision is made on the basis of the masked receptive fields Ω_r defined in (3). When reconstructing $\mathbf{v} \in \mathcal{V}$ from neuron \mathbf{r} , an error $\theta = \mathbf{v} - \mathbf{w}_r$ remains decomposed along the different directions, which results from projecting the output space grid back onto the input space \mathcal{V} :

$$\theta = \mathbf{v} - \mathbf{w}_r = \sum_{i=1}^{D_{\mathcal{A}}} a_i(\mathbf{v}) \frac{\mathbf{w}_{r+\mathbf{e}_i} - \mathbf{w}_{r-\mathbf{e}_i}}{\|\mathbf{w}_{r+\mathbf{e}_i} - \mathbf{w}_{r-\mathbf{e}_i}\|} + \mathbf{v}' \quad (10)$$

³ In the WEBSOM-application of KASKI the dimension was reduced from 5781 to approximately 150 [23].

Here, \mathbf{e}_i denotes the unit vector in direction i of \mathcal{A} and $a_i(\mathbf{v})$ are the projection amplitudes.⁴ Considering a receptive field Ω_r and determining its main (first) principal component $\boldsymbol{\omega}_{PCA}$ allows a further decomposition of \mathbf{v}' . Projection of \mathbf{v}' onto the direction of $\boldsymbol{\omega}_{PCA}$ then yields $a_{D_{\mathcal{A}}+1}(\mathbf{v})$,

$$\mathbf{v}' = a_{D_{\mathcal{A}}+1}(\mathbf{v}) \frac{\boldsymbol{\omega}_{PCA}}{\|\boldsymbol{\omega}_{PCA}\|} + \mathbf{v}'' \quad (11)$$

The *criterion for the growing* now is to add nodes in that direction which has on average the largest (normalized) expected error amplitude \tilde{a}_i :

$$\tilde{a}_i = \sqrt{\frac{n_i}{n_i + 1}} \sum_{\mathbf{v}} \frac{|a_i(\mathbf{v})|}{\sqrt{\sum_{j=1}^{D_{\mathcal{A}}+1} a_j^2(\mathbf{v})}}, i = 1, \dots, D_{\mathcal{A}} + 1 \quad (12)$$

After each growth step, a new learning phase has to take place in order to readjust the map. For a detailed study of the algorithm we refer to [5].

2.2 Magnification Control in SOMs

A further extension of the basic SOM concerns the so-called magnification. The standard SOM distributes the pointers $\mathbf{W} = \{\mathbf{w}_r\}$ according to the input distribution

$$\mathcal{P}(\mathbf{W}) \sim \mathcal{P}(\mathcal{V})^\alpha \quad (13)$$

with the magnification factor $\alpha = \frac{2}{3}$ [37], [25].⁵ The first approach to influence the magnification of a learning vector quantizer, proposed in [12] is called the *mechanism of conscience*. For this purpose a bias term is added in the winner rule (1):

$$\Psi_{\mathcal{V} \rightarrow \mathcal{A}} : \mathbf{v} \mapsto s(\mathbf{v}) = \underset{i \in \mathcal{A}}{\operatorname{argmin}} \left(\|\mathbf{v} - \mathbf{w}_i\| - \gamma \left(\frac{1}{N} - p_i \right) \right) \quad (14)$$

where p_i is the actual winning probability of the neuron i and γ is a balance factor. Hence, the winner determination is influenced by this modification. The algorithm should converge such that the winning probabilities of all neurons are equalized. This is related to a maximization of the entropy and consequently the resulting magnification is equal to unity.⁶ However, an arbitrary magnification cannot be achieved. Therefore, BAUER ET AL. in [3] introduced a local learning parameter ϵ_r

⁴ At the border of the output space grid, where not two, but just one neighboring neuron is available, we use $\frac{\mathbf{w}_r - \mathbf{w}_{r-\mathbf{e}_i}}{\|\mathbf{w}_r - \mathbf{w}_{r-\mathbf{e}_i}\|}$ or $\frac{\mathbf{w}_r + \mathbf{e}_i - \mathbf{w}_r}{\|\mathbf{w}_r - \mathbf{e}_i - \mathbf{w}_r\|}$ to compute the backprojection of the output space direction \mathbf{e}_i into the input space.

⁵ This result is valid for the one-dimensional case and higher dimensional ones which separate.

⁶ VAN HULLE points out that adding a conscience algorithm to the SOM does not equate to equiprobabilistic mapping, in general [21]. However, for *very high dimensions*, a minimum distortion quantizer (such as the conscience algorithm) approaches an equiprobable quantizer ([21] - page 93).

with $\langle \epsilon_{\mathbf{r}} \rangle \propto \mathcal{P}(\mathcal{V})^m$ in (4), where m is an additional control parameter. Equation (4) now reads as

$$\Delta \mathbf{w}_{\mathbf{r}} = \epsilon_s h_{\mathbf{r}\mathbf{s}} (\mathbf{v} - \mathbf{w}_{\mathbf{r}}) . \quad (15)$$

Note that the learning factor ϵ_s of the winning neuron s is applied to all updates. This local learning leads to a similar relation as in (13):

$$\mathcal{P}(\mathbf{W}) \sim \mathcal{P}(\mathcal{V})^{\alpha'} \quad (16)$$

with $\alpha' = \alpha(m+1)$ and allows a magnification control through the choice of m . In particular, one can achieve a resolution of $\alpha' = 1$, which maximizes mutual information [26,45].

2.3 Data Mining and Knowledge Discovery Using SOMs

If a proper SOM is trained according to the above mentioned criteria several methods for representation and post-processing can be applied. In case of a two-dimensional lattice of neurons many visualization approaches are known. The most common method for the visualization of SOMs is to project the weight vectors in the first dimension of the space spanned by the principle components of the data and connecting these units to the respective nodes in the lattice that are neighbored [23]. However, if the shape of the SOM lattice is hypercubical there exist several more ways to visualize the properties of the map. Here we concentrate only on those that are of interest in the applications presented later in this chapter. An extensive overview can be found in [39].

One interesting evaluation is the so-called **U**-matrix introduced by ULTSCH ET AL. [38]. The elements $U_{\mathbf{r}\mathbf{r}'}$ are the distances between the respective weight vectors $\mathbf{w}_{\mathbf{r}}$ and $\mathbf{w}_{\mathbf{r}'}$ where \mathbf{r} and \mathbf{r}' are neighbored in \mathcal{A}

$$U_{\mathbf{r}\mathbf{r}'} = \|\mathbf{w}_{\mathbf{r}} - \mathbf{w}_{\mathbf{r}'}\| \quad (17)$$

U can be used to determine clusters within the weight vector set and, hence, within the data space. Assuming that the map \mathcal{M} is approximately topology preserving, large values of **U** indicate cluster boundaries. If the lattice is a two-dimensional array the **U**-matrix can easily be viewed and gives a powerful tool for cluster analysis.

Another visualization technique can be used if the lattice \mathcal{A} is three-dimensional. The data points then can be mapped onto neuron \mathbf{r} can be identified by the color combination *red*, *green* and *blue* assigned to the location \mathbf{r} . In this way we are able to assign a color to each data point according to equation (1) and similar colors will encode groups of input patterns that were mapped close to one another in the lattice \mathcal{A} [42]. It should be emphasized that for a proper interpretation of this color visualization, as well as for the analysis of the **U**-matrix, topology preservation of the map \mathcal{M} is a strict requirement. The topology preserving property of \mathcal{M} must be proven prior to any evaluation of the map.

If we regard the SOM as a preprocessing method, the data can be analyzed in the lower-dimensional neuron space defined by the GSOM-generated lattice. Beside the

above mentioned U-matrix approach, a cluster algorithm such as Ward-clustering may be applied, taking the neighborhood relations into account explicitly [44]. An additional counter-propagation layer can also be added to the SOM, to learn a classification task in a supervised manner. This approach is faster than a MLP learning in many applications [19].

3 Remote Sensing Spectral Images

Spectral images can formally be described as a matrix $\mathcal{S} = \mathbf{v}^{(x,y)}$, where $\mathbf{v}^{(x,y)} \in \mathfrak{R}^{D_{\mathcal{V}}}$ is the vector of spectral information associated with pixel location (x, y) . The elements $v_i^{(x,y)}$, $i = 1 \dots D_{\mathcal{V}}$ of spectrum $\mathbf{v}^{(x,y)}$ reflect the responses of a spectral sensor at a suite of wavelengths (see Fig. 1). The spectrum is a characteristic pattern that provides a clue to the surface material within the area defined by pixel (x, y) . The individual 2-dimensional image $\mathcal{S}_i = v_i^{(x,y)}$ at wavelength i is called the i th image band.

The data space \mathcal{V} spanned by Visible-Near Infrared reflectance spectra is $[0 - \textit{noise}, U + \textit{noise}]^{D_{\mathcal{V}}} \subseteq \mathfrak{R}^{D_{\mathcal{V}}}$ where $U > 0$ represents an upper limit of the measured scaled reflectivity and \textit{noise} is the maximum value of noise across all spectral channels and image pixels. The data density $\mathcal{P}(\mathcal{V})$ may vary strongly within this space. Sections of the data 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. According to this model traditional multi-spectral imagery has a low $D_{\mathcal{V}}$ value while $D_{\mathcal{V}}$ can be several hundred for hyperspectral images. The latter case is of particular interest because the great spectral detail, complexity, and very large data volume pose new challenges in clustering, cluster visualization, and classification of images with such high spectral dimensionality [29].

In addition to dimensionality and volume, other factors, specific to remote sensing, can make the analyses of hyperspectral images even harder. For example, given the richness of data, the goal is to separate many cover classes, however, surface materials that are significantly different for an application may be distinguished by very subtle differences in their spectral patterns. The pixels can be mixed, which means that several different materials may contribute to the spectral signature associated with one pixel. Training data may be scarce for some classes, and classes may be represented very unevenly.

Noise is far less problematic than the intricacy of the 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.

3.1 Low-Dimensional Data: LANDSAT TM Multi-spectral Images

LANDSAT-TM satellite-based sensors produce images of the Earth in 7 different spectral bands. The ground resolution in meters is 30×30 for bands 1 – 5 and band

7. Band 6 (thermal band) has a spatial resolution of 60×60 only and it is often dropped from analyses. The LANDSAT TM bands were strategically determined for optimal detection and discrimination of vegetation, water, rock formations and cultural features within the limits of broad band multi-spectral imaging. The spectral information, associated with each pixel of a LANDSAT scene is represented by a vector $\mathbf{v} \in \mathcal{V} \subseteq \mathfrak{R}^{D_{\mathcal{V}}}$ with $D_{\mathcal{V}} = 6$. The aim of any classification algorithm is to subdivide this data space into subsets of data points, with each subset corresponding to specific features such as wood, industrial region, etc. The feature categories are specified by prototype data vectors (training spectra).

In the present contribution we consider two LANDSAT TM images. The first one is the north-east region of the city Leipzig in Germany⁷. The second one is from the Colorado area, U.S.A.⁸ For the Colorado image we also have a manually generated label map (ground truth image) for comparison. The labels indicate several regions of different vegetation.

A Grassberger-Procaccia analysis [15] of the Leipzig image yields $D^{\mathcal{GP}} \approx 1.7$ as an estimation for the intrinsic spectral dimension. The customary Principal Component Analysis results in the following vector of eigenvalues:

$$\mathbf{e}\mathbf{v} = (274.06, 76.19, 39.78, 11.92, 8.27, 6.28)^T \quad (18)$$

Application of the same two procedures to the Colorado image yields $D^{\mathcal{GP}} \approx 3.1414$ and

$$\mathbf{e}\mathbf{v} = (4.93, 0.68, 0.29, 0.05, 0.02, 0.02)^T, \quad (19)$$

respectively.

3.2 Hyperspectral Data: The Lunar Crater Volcanic Field AVIRIS Image

A Visible-Near Infrared ($0.4 - 2.5 \mu\text{m}$), 224-band, 20 m/pixel AVIRIS image of the Lunar Crater Volcanic Field (LCVF), Nevada, U.S.A., was analyzed in order to study SOM performance for high-dimensional remote sensing spectral imagery. (AVIRIS is the Airborne Visible-Near Infrared Imaging Spectrometer, developed at NASA/Jet Propulsion Laboratory. See <http://makalu.jpl.nasa.gov> for details on this sensor and on imaging spectroscopy.) The LCVF is one of NASA's remote sensing test sites, where images are obtained regularly. A great amount of accumulated ground truth from comprehensive field studies [2] and research results from independent earlier work such as [13] provide a detailed basis for the evaluation of the results presented here.

Fig. 2 shows a natural color composite of the LCVF with labels marking the locations of 23 different surface cover types of interest. This $10 \times 12 \text{ km}^2$ 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

⁷ obtained from UMWELT-FORSCHUNGSZENTRUM Halle-Leipzig, Germany

⁸ Thanks to M. Augusteijn (University of Colorado) for providing this image.

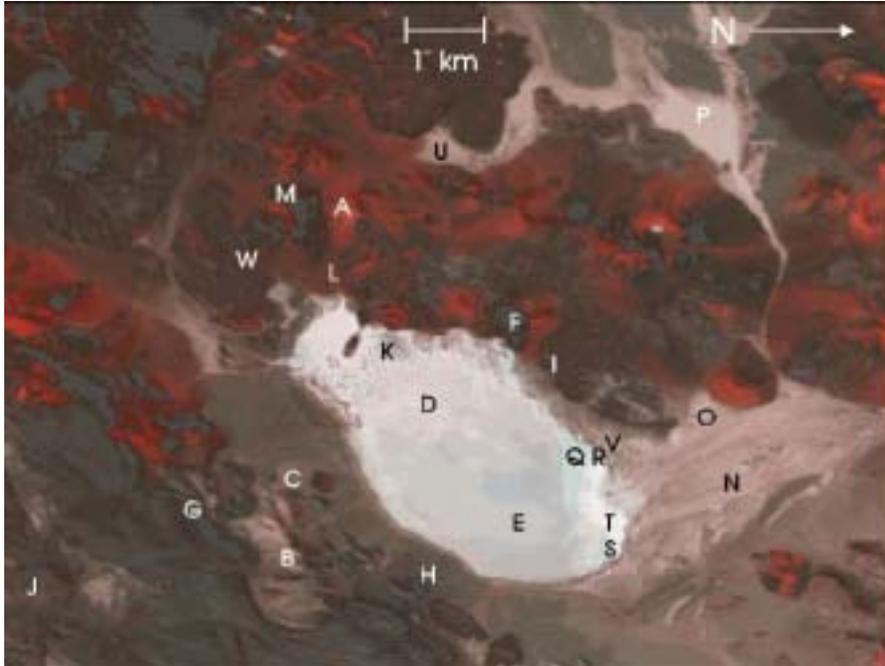


Fig. 2. The Lunar Crater Volcanic Field. RGB natural color composite from an AVIRIS, 1994 image. The original image comprises 224 image bands over the 0.4 - 2.5 μm range, 512 x 614 pixels, altogether 140 Mbytes of data. Labels indicate different cover types described in the text. The ground resolution is 20 m/pixel.

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 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 [29]). A long, NW-SE trending scarp, straddled by the label G, borders the vegetated area. Since this color composite only contains information from three selected image bands (one Red, one Green, and one Blue), many of the cover type variations remain undistinguished. They will become evident in the cluster and class maps below.

After atmospheric correction and removal of excessively noisy bands (saturated water bands and overlapping detector channels), 194 image bands remained from the original 224. These 194-dimensional spectra are the input patterns in the following analyses.

The spectral dimensionality of hyperspectral images is not well understood and it is an area of active research. While 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. Few investigations exist into the intrinsic dimensionality (ID) of hyperspectral images. Linear methods such as PCA or determination of mixture model endmembers [1] [22] usually yield 3 – 8 “endmembers”. BRUSKE [8] finds the spectral ID of the LCVF AVIRIS image (Fig. 2) to be between 3 and 7, using a non-linear neural network based approach (Optimally Topology Preserving Maps), whereas the Grassberger-Procaccia analysis [15] estimates the intrinsic dimension as $D^{GP} \approx 3.06$. These surprisingly low numbers, that increase with improved sensor performance [17], 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. With an associative neural network, Pendock [35] extracted 20 linear mixing endmembers from a 50-band ($2.0 - 2.5 \mu\text{m}$) segment of an AVIRIS image of Cuprite, Nevada (another well-known remote sensing test site), setting only a rather general surface texture criterium. Benediktsson et al. [6] performed feature extraction on an AVIRIS geologic scene of Iceland, 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 the same 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 chapter it is demonstrated that selection of 30 bands from the LCVF image in Fig. 2 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 [34] 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, [7] 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 tailor the data to available tools. An alternative is to develop capabilities to handle the full spectral information. Analysis of unreduced data is important for the establishment of benchmarks, 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.

4 SOM-Applications

4.1 Analysis of LANDSAT TM Images

One way to get good results for visualization of the clusters of LANDSAT TM data is to use a SOM dimension $D_{\mathcal{A}} = 3$ [18] and interpret the positions of the neurons \mathbf{r} in the lattice \mathcal{A} as vectors $\mathbf{r} = \mathbf{c} = (r, g, b)$ in the color space \mathcal{C} , where r, g, b are the intensities of the colors red, green and blue, respectively [18]. Such assignment of colors to winner neurons immediately yields a pseudo-color cluster map of the original image for visual interpretation. Since we are mapping the data clouds from a 6-dimensional input space onto a three-dimensional color space dimensional conflicts may arise and the visual interpretation may fail. However, in the case of topologically faithful mapping this color representation, prepared using all six LANDSAT TM image bands, contains considerably more information than a customary color composite combining three TM bands (frequently bands 2, 3, and 3). [18].

In the first LANDSAT example we investigate an image of the north-east region of Leipzig (described in sec. 3.1), using the GSOM approach. For comparison we also trained several regular SOMs with fixed output spaces the dimension of which range from 1 to 4, and determined the respective \tilde{P} -values, shown in Table 1. The topographic product \tilde{P} favors an output space dimension $D_{\mathcal{A}}$ between

N_{\max}	lattice structure	\tilde{P}
256	256	-0.189 ± 0.00612
256	16×16	-0.0642 ± 0.00031
252	$7 \times 6 \times 6$	$+0.0282 \pm 0.00024$
256	$4 \times 4 \times 4 \times 4$	$+0.0816 \pm 0.00387$

Table 1. Values of the topographic product \tilde{P} using different, fixed output spaces for the LANDSAT satellite image of Leipzig. For each structure, 3 or more runs were averaged.

2 and 3. However a clear decision cannot be made between 2 and 3. Yet, the edge

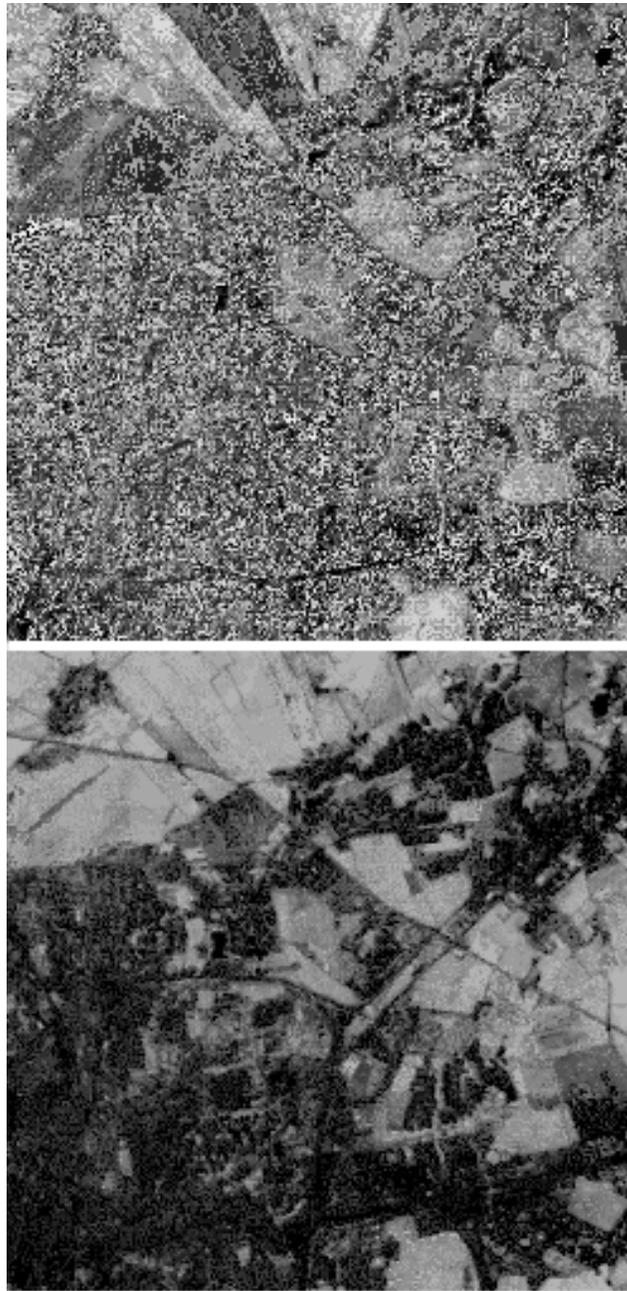


Fig. 3. Pseudo-color cluster maps of LANDSAT-TM six-band spectral images. Clusters of the Leipzig image using the $7 \times 6 \times 6$ standard SOM (top) and the $14 \times 6 \times 3$ GSOM-solution (bottom). (Color version available on request from villmann@informatik.uni-leipzig.de)

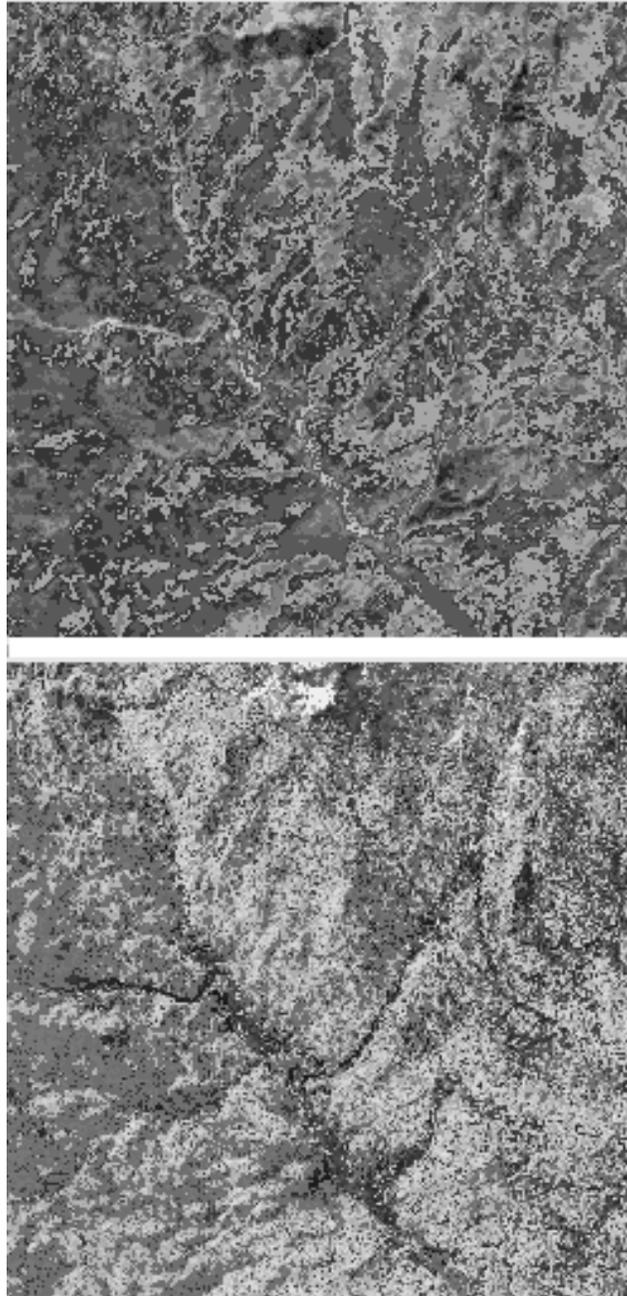


Fig. 4. Cluster maps of the Colorado image. The RGB color composite using bands 2, 3, and 4 (top), and the $12 \times 7 \times 3$ GSOM-solution derived from all six bands. (Color version available on request from villmann@informatik.uni-leipzig.de)

length ratios of the lattice provide a further free choice. We should emphasize that in these runs, as well as in the further GSOM simulations, we applied the new learning rule (15) to achieve maximal mutual information as pointed out in sec. 2.2. The GSOM algorithm was applied in several runs with different values of N_{\max} (maximum number of neurons allowed). The results obtained are depicted in Table 2. The

N_{\max}	lattice structure	\tilde{P}
128	$12 \times 5 \times 2$	0.0047
256	$14 \times 6 \times 3$	0.0050
512	$15 \times 6 \times 4$	0.0051

Table 2. GSOM results for the Leipzig image.

avored quasi two-dimensional structure is supported by the Grassberger-Procaccia analysis described above. Fig. 3 shows the visualization of the best GSOM solution with respect to the \tilde{P} -value. The \tilde{P} -values obtained by the GSOM are better than the respective values for the fixed lattice structures. Furthermore, for all values of N_{\max} , the GSOM yields approximately the same structure, and the length ratios of the edges are in a good agreement with the PCA eigenvalues (see (18)). However, in general, the standard linear PCA fails, as shown by the second LANDSAT image from the Colorado area. The PCA for this image suggests a one-dimensional structure (see (19)). The GSOM generates a $12 \times 7 \times 3$ lattice ($N_{\max} = 256$) in agreement with the Grassberger-Procaccia analysis ($D^{GP} \approx 3.1414$), which corresponds to a \tilde{P} -value of 0.0095 indicating good topology preservation (see Fig. 4).

4.2 SOM Analyses of Hyperspectral Imagery

A systematic supervised classification 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. The 23 geologically relevant classes indicated in Fig. 2 represent a great variety of surface covers in terms of spatial extent, the similarity of spectral signatures [29], and the number of available training samples. The full study, complete with evaluations of classification accuracies, is described in [31]. Average spectral shapes of these 23 classes are also shown in [29].

Fig. 5, top panel, shows the best classification, produced by an SOM-hybrid ANN using all 194 spectral bands that remained after preprocessing. 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. As mentioned



Fig. 5. Top: SOM-hybrid supervised 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 for a number of classes. Considerable loss of class distinction occurred compared to the ANN classification. 'bg' stands for background (unclassified pixels).

in sect. 2.3 such hybrid SOM constructions can also help faster learning. 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 [20] [32] [33]. Training samples for the supervised classifications 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 (SOM clusters), prior to training sample determination. Hence, Fig. 5 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. 5, 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.

Discovery in Hyperspectral Images with SOMs The previous section demonstrated the power of the SOM in helping discriminate among a large number of predetermined surface cover classes with subtle differences in the spectral patterns, using the full spectral resolution. It is even more interesting to examine the SOM’s performance in terms of detection of clusters in high-dimensional data. Fig. 6 displays a 40×40 SOM extended by the conscience algorithm of DeSieno according to equation (14). 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 input data, are indicated by various colors. The boundaries of these clusters were determined by a somewhat modified version of the U-matrix described in (17). 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. 6, are shown in the same colors in Fig. 7. Color coding for clusters that correspond to classes or subclasses of those in Fig. 5, top, is the same as in Fig. 5, to show similarities. Colors for additional groups were added.

The first observation is the striking correspondence between the supervised ANN class map in Fig. 5, 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 alert for 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. 5. 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. 5) 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. 5. 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). Evaluation of further refinements are left to the reader. Evidence that the SOM mapping in Fig. 6. approximates an equiprobabilistic mapping (that the magnification factor for the SOM in Fig. 6. is close to 1), using DeSieno's algorithm, is presented in [28].

As mentioned above, earlier investigations showed the intrinsic spectral dimensionality of the LCVF data set in the range of 3 – 7 [8]. The Grassberger-Procaccia analysis [15] yields $D_A^{GP} \approx 3.06$ corroborating the above results, i.e. the data are highly correlated, therefore a drastic dimensionality reduction may be possible. However, a faithful mapping is necessary to preserve the information contained in the hyperspectral image. For this purpose, the magnification control and the growing SOM (GSOM) procedure, as extensions of the standard SOM, are suitable tools. The GSOM produced a lattice of dimensions $8 \times 6 \times 6$, a radical dimension reduction. This is in agreement with the Grassberger-Procaccia analysis above. The resulting false color visualization of the spectral clusters is depicted in Fig. 8. It shows approximately the same quality as the $2d$ -SOM vector quantized but manually labeled image (supervised classification) in Fig. 5, top panel.

5 Conclusion

Self-Organizing Maps have been showing great promise for the analyses of remote sensing spectral images. With recent advances in remote sensor technology, very high-dimensional spectral data emerged and demand new and advanced approaches to cluster detection, visualization, and supervised classification. While standard SOMs produce good results, the high dimensionality and large amount of hyperspectral data call for very careful evaluation and control of the faithfulness of topological mapping performed by SOMs. Faithful topological mapping is required in order to avoid false interpretations of cluster maps created by an SOM. This chapter summarized several advances that were made in the past few years, and that

ensure strict topology preservation through mathematical considerations. Two of these extensions to the standard Kohonen SOM, the Growing Self-Organizing Map and magnification control, were discussed in detail, along with their relationship to other SOM extensions such as the DeSieno conscience mechanism, or to independent analyses such as the Grassberg-Procaccia analysis for the determination of intrinsic dimensionality. Case studies on real multi- and hyperspectral images were presented that support our theoretical discussions. While it is outside the scope of this contribution, as a final note we want to point out that full scale investigations such the LCVF study in this chapter also have to make heavy use of advanced image processing tools and user interfaces, to handle great volumes of data efficiently, and for effective graphics/visualization. References to such tools are made in the cited literature on data analyses.

Acknowledgements

E.M. has been supported by the Applied Information Systems Research Program of NASA, Office of Space Science, NAG54001 and NAG59045. Contributions by Dr. William H. Farrand, providing field knowledge and data for the evaluation of the LCVF results are gratefully acknowledged.

References

1. J. B. Adams, M. O. Smith, and A. R. Gillespie. Imaging spectroscopy: Interpretation based on spectral mixture analysis. In C. Peters and P. Englert, editors, *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pages 145–166. Cambridge University Press, New York, 1993.
2. R. E. Arvidson and M. D.-B. and. et al. Archiving and distribution of geologic remote sensing field experiment data. *EOS, Transactions of the American Geophysical Union*, 72(17):176, 1991.
3. H.-U. Bauer, R. Der, and M. Herrmann. Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, 8(4):757–771, 1996.
4. H.-U. Bauer and K. R. Pawelzik. Quantifying the neighborhood preservation of Self-Organizing Feature Maps. *IEEE Trans. on Neural Networks*, 3(4):570–579, 1992.
5. H.-U. Bauer and T. Villmann. Growing a Hypercubical Output Space in a Self-Organizing Feature Map. *IEEE Transactions on Neural Networks*, 8(2):218–226, 1997.
6. J. A. Benediktsson, J. R. Sveinsson, and et al. Classification of very-high-dimensional data with geological applications. In *Proc. MAC Europe 91*, pages 13–18, Lenggries, Germany, 1994.
7. J. A. Benediktsson, P. H. Swain, and et al. Classification of very high dimensional data using neural networks. In *IGARSS'90 10th Annual International Geoscience and Remote Sensing Symp.*, volume 2, page 1269, 1990.
8. J. Bruske and E. Merényi. Estimating the intrinsic dimensionality of hyperspectral images. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'99)*, pages 105–110, Brussels, Belgium, 1999. D facto publications.
9. J. Campbell. *Introduction to Remote Sensing*. The Guilford Press, U.S.A., 1996.
10. R. N. Clark. Spectroscopy of rocks and minerals, and principles of spectroscopy. In A. Rencz, editor, *Manual of Remote Sensing*. John Wiley and Sons, Inc, New York, 1999.
11. R. Der, G. Balzuweit, and M. Herrmann. Constructing principal manifolds in sparse data sets by self-organizing maps with self-regulating neighborhood width. In *ICNN 96. The 1996 IEEE International Conference on Neural Networks (Cat. No. 96CH35907)*, volume 1, pages 480–483. IEEE, New York, NY, USA, 1996.
12. D. DeSieno. Adding a conscience to competitive learning. In *Proc. ICNN'88, Int. Conf. on Neural Networks*, pages 117–124, Piscataway, NJ, 1988. IEEE Service Center.
13. W. H. Farrand. *VIS/NIR Reflectance Spectroscopy of Tuff Rings and Tuff Cones*. PhD thesis, University of Arizona, 1991.
14. J. Goppert and W. Rosenstiel. The continuous interpolating self-organizing map. *Neural Processing Letters*, 5(3):185–92, 1997.
15. P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica*, 9D:189–208, 1983.
16. R. O. Green. Summaries of the 6th Annual JPL Airborne Geoscience Workshop. 1. AVIRIS Workshop, Pasadena, CA, March 4–6 1996.
17. R. O. Green and J. Boardman. Exploration of the relationship between information content and signal-to-noise ratio and spatial resolution. In *Proc. 9th AVIRIS Earth Science and Applications Workshop*, Pasadena, CA, February, 23–25 2000.
18. M. H. Gross and F. Seibert. Visualization of multidimensional image data sets using a neural network. *Visual Computer*, 10:145–159, 1993.
19. J. A. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*, volume 1 of *Santa Fe Institute Studies in the Sciences of Complexity: Lecture Notes*. Addison-Wesley, Redwood City, CA, 1991.

20. E. S. Howell, E. Merényi, and L. A. Lebofsky. Classification of asteroid spectra using a neural network. *Jour. Geophys. Res.*, 99(10):847–865, 1994.
21. M. M. V. Hulle. *Faithful Representations and Topographic Maps*. Wiley Series and Adaptive Learning Systems for Signal Processing, Communications, and Control. Wiley Sons, New York, 2000.
22. R. S. Inc. *ENVI v.3 User's Guide*, 1997.
23. S. Kaski, J. Nikkilä, and T. Kohonen. Methods for interpreting a self-organized map in data analysis. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, pages 185–190, Brussels, Belgium, 1998. D facto publications.
24. T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).
25. T. Kohonen. Comparison of SOM point densities based on different criteria. *Neural Computation*, 11(8):212–234, 1999.
26. R. Linsker. How to generate maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1:402–411, 1989.
27. T. Martinetz and K. Schulten. Topology representing networks. *Neural Networks*, 7(3):507–522, 1994.
28. E. Merényi. "Precision mining" of high-dimensional patterns with self-organizing maps: Interpretation of hyperspectral images. In P. Sinčák and J. Vasčák, editors, *Quo Vadis Computational Intelligence? New Trends and Approaches in Computational Intelligence (Studies in Fuzziness and Soft Computing, Vol. 54)*. Physica-Verlag, 2000.
29. E. Merényi. Self-organizing ANNs for planetary surface composition research. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, pages 197–202, Brussels, Belgium, 1998. D facto publications.
30. E. Merényi. The challenges in spectral image analysis: An introduction and review of ANN approaches. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'99)*, pages 93–98, Brussels, Belgium, 1999. D facto publications.
31. E. Merényi, W. H. Farrand, and et al. Efficient geologic mapping from hyperspectral images with artificial neural networks classification: a comparison to conventional tools. *IEEE TGARS*, in preparation, 2001.
32. E. Merényi, E. S. Howell, and et al. Prediction of water in asteroids from spectral data shortward of 3 microns. *ICARUS*, 129(10):421–439, 1997.
33. 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*, 124(10):280–295, 1996.
34. T. Moon and E. Merényi. Classification of hyperspectral images using wavelet transforms and neural networks. In *Proc. of the Annual SPIE Conference*, page 2569, San diego, CA, 1995.
35. N. Pendock. A simple associative neural network for producing spatially homogeneous spectral abundance interpretations of hyperspectral imagery. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'99)*, pages 99–104, Brussels, Belgium, 1999. D facto publications.
36. H. Ritter. Parametrized self-organizing maps. In S. Gielen and B. Kappen, editors, *Proc. ICANN'93 Int. Conf. on Artificial Neural Networks*, pages 568–575, London, UK, 1993. Springer.
37. H. Ritter and K. Schulten. On the stationary state of Kohonen's self-organizing sensory mapping. *Biol. Cyb.*, 54:99–106, 1986.
38. A. Ultsch. Self organized feature maps for monitoring and knowledge acquisition of a chemical process. In S. Gielen and B. Kappen, editors, *Proc. ICANN'93, Int. Conf. on Artificial Neural Networks*, pages 864–867, London, UK, 1993. Springer.

39. J. Vesanto. SOM-based data visualization methods. *Intelligent Data Analysis*, 3(7):123–456, 1999.
40. T. Villmann. Benefits and limits of the self-organizing map and its variants in the area of satellite remote sensing processing. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'99)*, pages 111–116, Brussels, Belgium, 1999. D facto publications.
41. T. Villmann. Topology preservation in self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, number ISBN 951-22-3589-7, pages 279–292, Amsterdam (Holland), June 1999. Helsinki, Elsevier.
42. T. Villmann. Neural networks approaches in medicine – a review of actual developments. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2000)*, pages 165–176, Brussels, Belgium, 2000. D facto publications.
43. T. Villmann, R. Der, M. Herrmann, and T. Martinetz. Topology Preservation in Self-Organizing Feature Maps: Exact Definition and Measurement. *IEEE Transactions on Neural Networks*, 8(2):256–266, 1997.
44. T. Villmann, W. Hermann, and M. Geyer. Data mining and knowledge discovery in medical applications using self-organizing maps. In R. Brause and E. Hanisch, editors, *Medical Data Analysis*, pages 138–151, Berlin, New York, Heidelberg, 2000. Lecture Notes in Computer Science 1933, Springer-Verlag.
45. T. Villmann and M. Herrmann. Magnification control in neural maps. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, pages 191–196, Brussels, Belgium, 1998. D facto publications.



Fig. 6. Clusters identified in a 40×40 SOM. The SOM was trained on the entire 194-band LCVF image, using the DeSieno [12] algorithm.

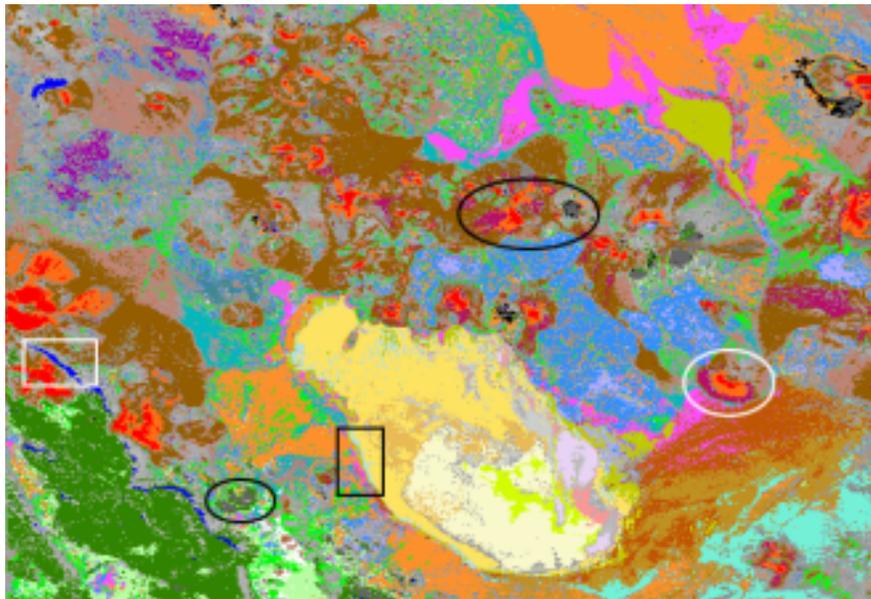


Fig. 7. The clusters from Fig.6 remapped to the original spatial image, to show where the different spectral types originated from. The relatively large, light grey areas correspond to the black, unevaluated parts of the SOM in Fig.6. Ovals and rectangles highlight examples discussed in the text.

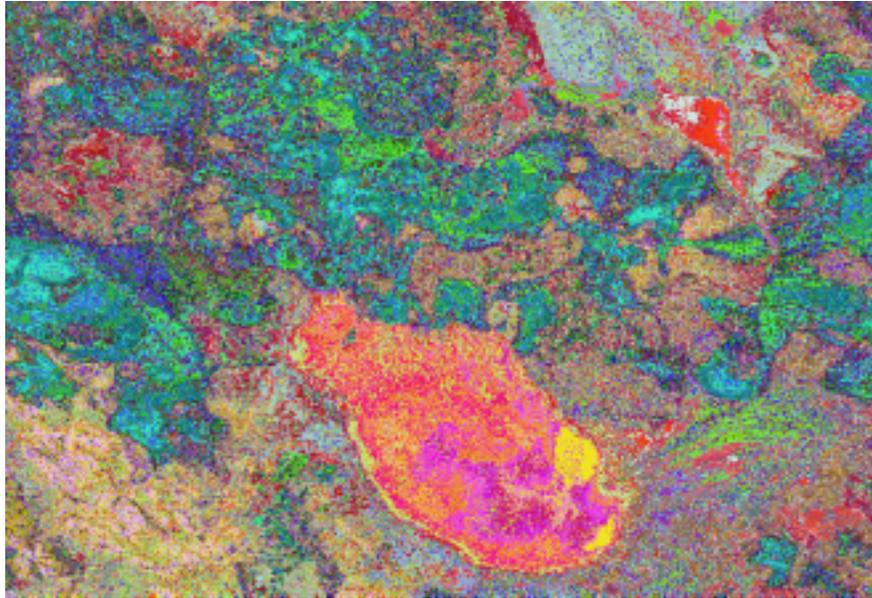


Fig. 8. GSOM generated false color cluster map of the same 194 band hyperspectral image of the Lunar Crater Volcanic Field, Nevada, USA, as in Fig.5. It shows similar groups as seen in the supervised classification map in Fig. 5, top panel.