Explicit Magnification Control of Self-Organizing Maps for "Forbidden" Data

Erzsébet Merényi, *Senior Member, IEEE*, Abha Jain, and Thomas Villmann

Abstract

We examine the scope of validity of the explicit SOM magnification control scheme of Bauer, Der, and Herrmann [1], on data for which the theory does not guarantee success, namely data that are n-dimensional, $n \geq 2$ and whose components in the different dimensions are not statistically independent. The Bauer *et al.* algorithm is very attractive for the possibility of faithful representation of the *pdf* of a data manifold, or for discovery of rare events, among other properties. Since theoretically unsupported data of higher dimensionality and higher complexity would benefit most from the power of explicit magnification control, we conduct systematic simulations on "forbidden" data. For the unsupported n = 2 cases that we investigate the simulations show that even though the magnification exponent $\alpha_{achieved}$ achieved by magnification control is not the same as the desired $\alpha_{desired}$, $\alpha_{achieved}$ systematically follows $\alpha_{desired}$ with a slowly increasing positive offset. We show that for simple synthetic higher-dimensional data information theoretically optimum *pdf* matching ($\alpha_{achieved} = 1$) can be achieved, and that negative magnification has the desired effect of improving the detectability of rare classes. In addition we further study theoretically unsupported cases with real data.

***Index Terms*— Self-Organizing Maps, map magnification, data mining, high-dimensional data.**

I. Potential benefits, and known limits of SOM magnification control

ONE theoretically interesting and powerful data analysis aspect of Self-Organizing Maps (SOMs) is the *map magnification*, which relates the density Q of SOM weights in the input space, to the *pdf* P of the input data by the following power law:

$$Q(\mathbf{w}) = cP(\mathbf{w})^{\alpha} \tag{1}$$

Here α is the *magnification exponent* and c is a constant [1]. Certain values of α have been associated with particular quantization or information theoretical properties [1], [2]. An SOM with $\alpha = 1$ maximizes information theoretic entropy, therefore such mapping produces the best approximation to the *pdf* of the data with the given number of codebook vectors (SOM weights). $\alpha = 1/3$ for 1-dimensional data corresponds to minimum mean squared error quantization, or, in general, $\alpha = d/(d + p)$ leads to minimum mean squared error quantization of d-dimensional data in p-norm [2]. Thus magnification control enables the realization of different similarity concepts in the underlying cost function (description error) based on the respective p-norms. For example, near-zero values for α approximate the maximum norm in the description error. As suggested in [1] on the basis of biological observations of the "perceptual magnet" effect that enlarges the cortical representation areas for rare events (e.g., events of particular danger that an organism must be alert to), $alpha < 0$ can enable better categorization by allocating larger-than-proportional areas of the SOM lattice for low-frequency inputs. Regions of lower data probability become more accurately represented by prototypes, which can lead to detection of rare clusters. That, in turn, increases the chance of discovery. In data mining, this mechanism could alert for very small, "interesting" or "suspicious" groupings in data such as caused by mineralogical deposits of tiny spatial extent that may be indication of past life in Mars imagery, or by terrorist activity in data gathered by security agencies. As is known, the inherent property of the Kohonen SOM (KSOM) [3] is a map magnification of $\alpha = 2/3$ (under certain conditions) [4], [5]. This value of α is optimal in neither minimum distortion nor maximum entropy sense. A SOM variant called Conscience algorithm [6] is constructed to achieve $\alpha = 1$, but cannot induce any other value. The Conscience algorithm is based on heuristics, and although it works well in practice, theoretical proof does not exist for the achieved map magnification.

Controlling the magnification of Self-Organizing Neural Maps is therefore an extremely attractive possibility because various values of the magnification exponent can effect desirable quantization properties and serve specific data mining purposes.

The explicit SOM magnification control introduced by Bauer, Der, and Herrmann [1] (referred to as BDH from now on) provided a powerful principled approach to obtaining a desired magnification exponent for 1-dimensional data and for n-dimensional data whose components are statistically independent. To briefly quote the method, let V and A denote the input data manifold and the SOM grid of Processing Elements (PEs, nodes, or neurons), respectively, P and Q the *pdf* of V and the *pdf* of the SOM weights in V, respectively. The SOM PEs are indexed by their (potentially multi-dimensional) grid

Erzsébet Merényi is at Rice University, Electrical and Computer Engineering Department, e-mail: erzsebet@rice.edu

Abha Jain was at Rice University at the time of this work, now she is with Microsoft Corporation, Redmond

Thomas Villmann is at University of Leipzig, Clinic for Psychotherapy Leipzig, Germany, e-mail: thomas.villmann@medizin.uni-leipzig.de

Figures in this paper are in color. Request color reprint from erzsebet@rice.edu.

locations $\mathbf{r}$. The weight attached to node $\mathbf{r} \in A$ is $\mathbf{w_r}$. For any $\mathbf{v} \in V$ input the Kohonen SOM learning algorithm [7] selects a winner node $\mathbf{s}$ by

$$\mathbf{s} = \underset{\mathbf{r}}{argmin}||\mathbf{v} - \mathbf{w_r}|| \tag{2}$$

and then updates the weights according to

$$\Delta \mathbf{w_r} = \epsilon h_{\mathbf{rs}}(\mathbf{v} - \mathbf{w_r}) \tag{3}$$

where the *neighborhood function* $h_{\mathbf{rs}}$ defines which weights get updated. $h_{\mathbf{rs}}$ is typically a Gaussian function centered over the winner node, which we will assume here. In general, the neighborhood can be defined many different ways. In equation (3) the learning rate ϵ is globally defined, *i.e.*, it is the same for all PEs for a given time step and its value is independent of any local properties of the map. The key idea of the BDH learning is to make the learning rate dependent on the local input density so as to effect

$$\epsilon_{\mathbf{r}} \propto \epsilon_0 P(\mathbf{w_r}) \tag{4}$$

This goal is achieved by changing ϵ in equation (3) to

$$\epsilon_{\mathbf{s}}(t) = \epsilon_0 \left(\frac{1}{\Delta t_{\mathbf{s}}} \left(\frac{1}{||\mathbf{v} - \mathbf{w_s}||^d} \right) \right)^m \tag{5}$$

where m is a free parameter, t is the time step and $\Delta t_{\mathbf{s}}$ is the time difference since the PE $\mathbf{s}$ won last. ϵ_0 is a constant, and d denotes the effective dimensionality of the receptive field of $\mathbf{w_s}$. Note that $\epsilon_{\mathbf{s}}(t)$ is locally determined but then applied to *all* weight updates in the current step. By doing so the local property of the map is propagated to the neighbors. With a derivation similar to that of Ritter and Schulten [4] in which the $\alpha = 2/3$ property of the KSOM was proven, Bauer *et al.* showed that the learning rate in equation (5) modifies the power law under equation (1) to

$$Q(\mathbf{w}) = cP(\mathbf{w})^{\alpha'} = P(\mathbf{w})^{\frac{2}{3}(1+m)} \tag{6}$$

where the free parameter m can be used for controlling the value of α'. Therefore, to achieve a desired magnification exponent α', $m = 3/2 * \alpha' - 1$ needs to be used in equation (5). The reader is referred to [1] for further details.

Most real data, of course, do not obey the conditions stipulated for BDH learning (listed in Section II below), yet it is real data scenarios that would benefit the most from explicit magnification control. Extension of analytical proof of the BDH scheme for such higher-dimensional cases hinges on analytical proof of the SOM as seen from, *e.g.*, [4], with which there has not been much luck so far. This, however, does not necessarily mean that the BDH algorithm should not work for more complex data than it is proven for.

This work examines and extends the limits of the BDH, through carefully constructed numerical simulations. It also extends and makes more precise our preliminary results published in [8], [9].

II. The behavior of the BDH scheme on 1- and 2-dimensional data

The theory put forward by Bauer *et al.* [1] proves that the BDH algorithm will successfully induce the intended value of α for the following cases:

- 1-dimensional input data
- n-dimensional data, $\mathbf{v} = (v_1, \ldots v_n)$, if and only if $p_{\mathbf{v}}(\mathbf{v}) = p_{v_1}(v_1) * \ldots p_{v_n}(v_n)$ (*i.e.*, the *pdf* factorizes into the marginals)

We examine the possible validity of the BDH algorithm beyond these limited cases. We specify various values of the magnification exponent α as the desired target value to be induced by the BDH. In each case, α achieved by the converged SOM is calculated by a histogram based method as used in [10] and compared to the desired α. Details of the histogram method are discussed in [11]. We provide a basic description in the Appendix after the Discussion section.

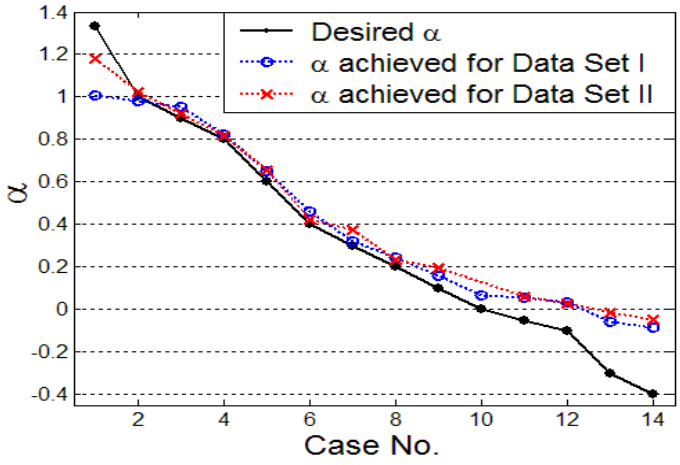

Fig. 1. Results of magnification control on 1-dimensional data for Data Set I (blue open circles) and Data Set II (red x-es). The graphs compare the values of the magnification exponent, α, obtained through the explicit magnification control scheme of Bauer *et al.* [1] with the desired values of α (black dots) that were input to the BDH algorithm. The obtained values were derived by the evaluation of the converged SOM, similarly as described in [10], using the known *pdf*s of Data Sets I and II. Case No. refers to separate runs of the BDH algorithm with different desired α values.

First we confirm that the BDH works well for the above, theoretically supported cases. Simulations with 1-dimensional data, similar to those in [1], are shown in Figure 1 for two 1-dimensional data sets. Data Set I is generated by $p(v) = 2v, v \in [0, 1]$ and Data Set II is generated by $p(v) = 3v^2, v \in [0, 1]$. As seen from Figure 1 the values of the magnification exponent to be induced by the BDH are in tight correspondence with the actual values achieved, between $0 \leq \alpha_{desired} \leq 1$. The slight discrepancies can be attributed to the granularity of the quantization (100 SOM weights). The discrepancies are appreciable for $\alpha_{desired} < 0$ and for $\alpha_{desired} > 1$, however, it is worth noting that the deviations are not random: the $\alpha_{achieved}$ values seem to be attracted to 1 or 0 (for $\alpha_{desired} > 1$ and $\alpha_{desired} < 0$), respectively. The source of the obvious discrepancies outside of the $0 \leq \alpha_{desired} \leq 1$ range could be due to less accuracy in

the calculation of $\alpha_{achieved}$ or instability of the magnification control. To our knowledge, this has not been investigated for the BDH. We know, however, from analyses for other magnification control approaches that instability regions may exists [12].

Next we investigate the performance of the BDH algorithm on simple 2-dimensional data. The experiment **a)** will verify the BDH under the stipulated theoretical conditions of independency. Experiments **b)** and **c)** explore theoretically unsupported cases where data in the two dimensions are weakly and strongly correlated, respectively. The following cases, with different correlations (ρ) between the two dimensions of the data space, are evaluated:

a) Data independent in the two dimensions: $p_{\mathbf{v}}(\mathbf{v}) = p_{v_1}(v_1)p_{v_2}(v_2)$, generated according to the following *pdf*:

$$\begin{aligned} p_{v_1} &= 2v_1, \quad v_1 \in [0,1] \\ p_{v_2} &= 2v_2, \quad v_2 \in [0,1] \\ \mathrm{p}_{\mathbf{v}}(\mathbf{v}) &= 4v_1v_2, \quad v_1, v_2 \in [0,1] \end{aligned} \tag{7}$$

As demonstrated in Figure 2 the BDH magnification control works as advertised, for this data set.

b) Data weakly correlated in two dimensions: $\rho \ll 1$. The data consists of two equal size subsets of 2-dimensional samples, defined by

$$\begin{aligned} \mathbf{v} &= (v_1, v_2) \\ v_2 &= v_1 + n \quad \text{for one subset} \\ v_2 &= -v_1 + n \quad \text{for the second subset} \end{aligned} \tag{8}$$

where $n = \mathcal{N}(0, 0.0625)$. v_1 and v_2 are weakly correlated with the correlation coefficient $\rho_{v_1v_2} = 0.0044$. From Figure 3 it can be seen that the α achieved and the desired α are almost equal at $\alpha = 1$ and the two values differ increasingly but in a predictable manner as α decreases. This is a stronger result than available from the theory, as the theory only guarantees successful prediction if and only if v_1 and v_2 are independent.

c) Data strongly correlated in two dimensions: $\rho \approx 1$. This data set consists of 2-dimensional samples generated by

$$\begin{aligned} \mathbf{v} &= (v_1, v_2) \\ v_2 &= v_1 + n \end{aligned} \tag{9}$$

where $n = \mathcal{N}(0, 0.25)$. The correlation coefficient is $\rho_{v_1v_2} = 0.9026$. In this strongly correlated case, even though α achieved by the map differs from the desired α value, there is a clearly observable trend that the achieved values of α are systematically decreasing, following the desired values with a shift that increases slowly with decreasing α values (Figure 4). This is again a stronger result than the theory provides.

Additional 1- and 2-dimensional cases are analyzed and more details given on analysis considerations in [11].

III. BDH ON HIGHER-DIMENSIONAL DATA

The BDH algorithm has no analytical justification for data with 2 or more non-separating dimensions. However, the systematic results from numerical simulations on 2-dimensional "forbidden" data in the previous section encourage investigation of the BDH behavior on higher-dimensional data.

Our methodology to examine the performance of the BDH algorithm on 1- and 2-dimensional data sets was to induce a certain value of α, evaluate the α achieved by the converged SOM, and compare the two. However, evaluation of α is not an easy task in general, especially when the input *pdf* is unknown, as is most commonly the case. The evaluation of α involves the estimation of the *pdf* of the data and that of the SOM weights. So far, we were using a histogram based method for α evaluation [10]. This method becomes inapplicable for high-dimensional data as the number of samples required for the estimation of the *pdf* increases exponentially with dimensionality. For the 6- and 8-dimensional data in this study the required number of samples for *pdf* estimation could easily be over 10 million, a number that far exceeds the total number of our data samples in each case. A summary of requirements of *pdf* estimation for higher-dimensional data is given in [11]. Here we have to restrict ourselves to giving a brief introduction, along with further pointers to related issues and references in the Appendix.

There exist several neural approaches to estimate the *pdf* of the data on the prototype level. Prototype based approaches are the DeSieno Conscience Learning [6], Frequency-Sensitive Competitive Learning [13], the magnification controlled neural gas [14], kernelized variants of the SOM [15] and a Gaussian mixture approach based on SOM [16], to name just a few. However, comparison could be difficult. For example, Van-Hulle reports, based on 1-dimensional experiments on artificial data, that Conscience Learning could be unstable for precise *pdf* estimation because the conscience may be taken up by only a few prototypes [15]. Also, there is no theoretical proof for Conscience Learning. Frequency-Sensitive Competitive Learning is theoretically proven only for the one-dimensional case and we do not know of numerical verification for higher dimensions.

So for higher-dimensional cases, we evaluate the performance of the algorithm indirectly: by observing the resulting map and comparing to independently known properties of the input data. We concentrate on cases where we can evaluate the magnification results in some meaningful way, from an application's point of view. One special case is *pdf* estimation (forcing $\alpha = 1$), another is negative magnification. Both can be particularly useful for complex, high-dimensional data if systematic experiments indicate a *predictable* behavior of the BDH for such "forbidden data". For example, if successful forcing of negative magnification can be shown, the magnification effect is useful for data mining regardless of the exact value of the induced magnification exponent, which we cannot calculate.

A. BDH-induced negative magnification on 6-dimensional synthetic data

Finding very small classes in a data set is a challenging task. Input classes of rare occurrence may have little or no representation in the output map when KSOM is used. As

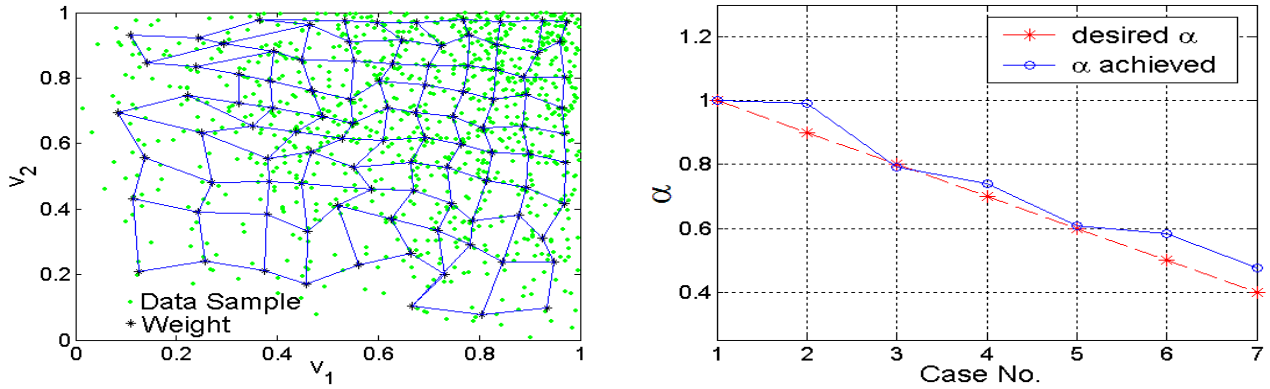

Fig. 2. BDH simulation with the data set defined under a), whose 2-dimensional input samples $\mathbf{v} = (v_1, v_2)$ are such that v_1 and v_2 are independent. **Left:** 2-dimensional input data samples (small green dots) and distribution of SOM weights (larger black dots), resulting from BDH with $\alpha_{desired} = 0.6$ after $2,000,000$ learning steps. Weights adjacent in the SOM lattice are connected. **Right:** Comparison of the achieved $\alpha_{achieved}$ to $\alpha_{desired}$ shows very good agreement, as expected. The discrepancies are largely due to the fact that the theoretical results are asymptotic and we only have a finite number of PEs (100).

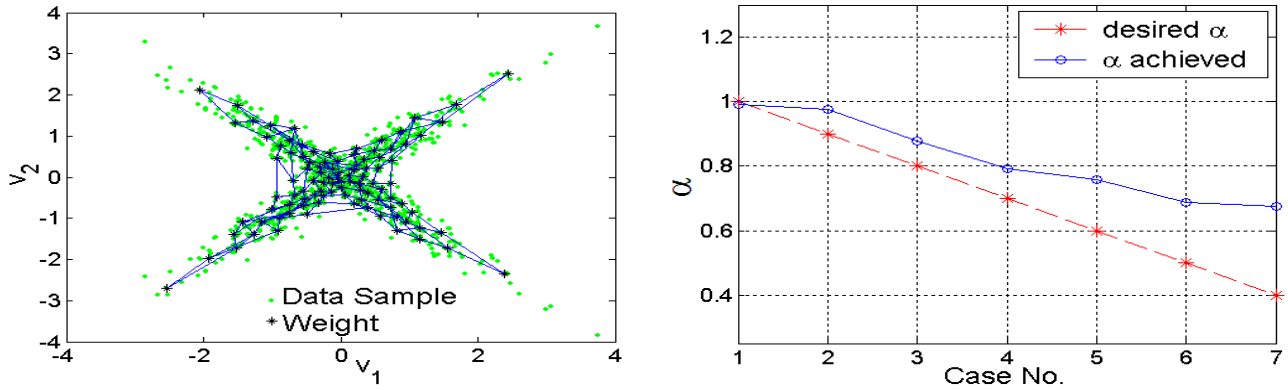

Fig. 3. BDH simulation with the data set under b), whose 2-dimensional input samples $\mathbf{v} = (v_1, v_2)$ are such that v_1 and v_2 are weekly correlated, $\rho_{v_1 v_2} \ll 1$. **Left:** 2-dimensional input data samples (small green dots) and distribution of SOM weights (larger black dots), resulting from BDH with $\alpha_{desired} = 1.0$ after $2,000,000$ steps. Weights adjacent in the SOM are connected. **Right:** The difference between $\alpha_{achieved}$ achieved and $\alpha_{desired}$ increases in a predictable manner as α decreases from 1.

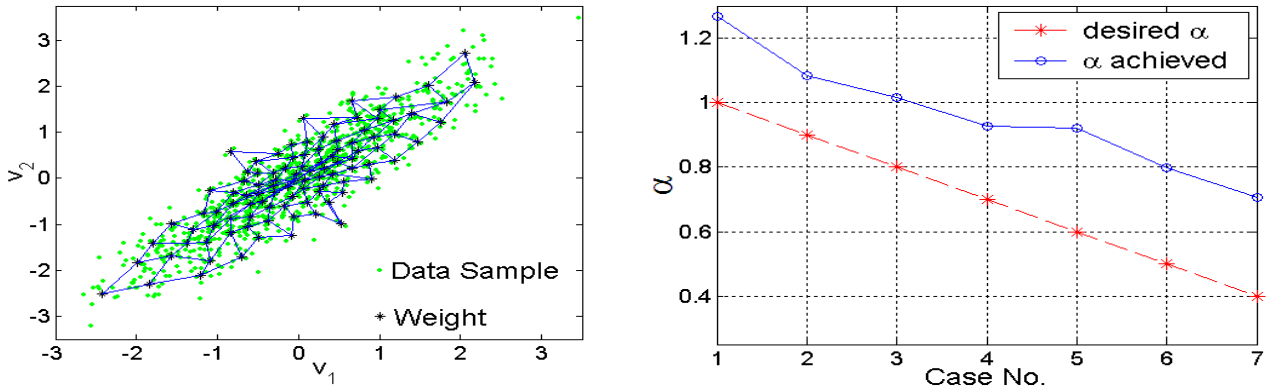

Fig. 4. BDH simulation with the data set under c), whose 2-dimensional input samples $\mathbf{v} = (v_1, v_2)$ are such that v_1 and v_2 are strongly correlated, $\rho_{v_1 v_2} \approx 1$. **Left:** 2-dimensional input data samples (small green dots) and distribution of SOM weights (larger black dots), resulting from BDH with $\alpha_{desired} = 1.0$ after $1,000,000$ steps. Weights adjacent in the SOM are connected. **Right:** The difference between $\alpha_{achieved}$ achieved and $\alpha_{desired}$ shows a clear trend of a shift that slowly increases with decreasing $\alpha_{desired}$.

shown for 2-dimensional factorizing data in [1], application of BDH with $\alpha < 0$ results in negative magnification: the areal representation of low-frequency input samples becomes enlarged in the SOM. This is promising for detection of rare classes. Does it, however, work for higher dimensions?

We use two synthetic 6-band (spectral) images with known cluster structure, described in Figure 5, to show that negative magnification can be induced on this type of "forbidden" data. A spectral image consists of n co-registered image bands, each of which is taken at a different wavelength. Every pixel of the image, therefore, is characterized by an n-dimensional vector, called the spectrum, similar to the data in Figure 5. In real spectral images, taken of material surfaces, the spectrum carries compositional information about the material in the respective pixel. See, for example, [17] for more detail. Spectral images are powerful information sources and are used in many areas of scientific research, business, industry, defense systems, etc. Detailed and precise exploitation of such data is of great interest. One especially valuable capability is the discovery of small, interesting groups of data.

The synthetic data sets used in this section are 128 x 128 pixel images, whith a 6-dimensional vector associated with each image pixel. These 6-dimensional vectors are the input patterns for the SOM. Both images are artificially divided into rectangular areas within which the spectral signatures are the same. The spatial distribution of the various classes, along with their mean spectral signatures are described in Figure 5. The two images represent two levels of data complexity. The 5-class data is illustrated on the left in Figure 5. Class U is a rare class with only 1 data point and with a spectral signature that is very different from the signatures of the other four classes. The rest of the classes have 4096 or 4095 data points each. In this data set, correlation coefficients between the different dimensions range from $0.004(\rho_{v_2v_3})$ to $0.9924(\rho_{v_3v_6})$, which renders it a "forbidden" case for application of BDH. When a 10 x 10 KSOM is used, the rare class U is represented by only one Processing Element in the SOM (Figure 6, left). BDH with $\alpha_{desired} = -0.8$ magnifies the rare class in the map: in Figure 6, right, it is represented by 10 PEs!

The choice of $\alpha_{desired} = -0.8$ comes from an extrapolation of the curves in Figure 4, from which we estimated that a value of $0 > \alpha_{desired} > -0.5$ will likely induce $\alpha_{achieved} \approx 0$, so in order to ensure a significant negative induced value we need to specify $\alpha_{desired} \ll -0.5$. On the other end, Figure 1 cautions us not to choose too large negative values. This extrapolation is admittedly only a hypothesis at present, not only because the simulations in Figure 4 were not conducted for negative values but also because even if we had values for 2-dimensional data assuming the same behavior for higher-dimensional data would still remain a hypothesis. Evidence of successful forcing of negative magnification may motivate extension of the Figure 4 simulations to the negative range in future work.

The synthetic image data set with 20 classes is similar to the 5-class image data set in concept, but it has 20 classes, as shown in Figure 5, right. Two of the classes, marked R and Q are very small, with only 1 and 16 data points, respectively. Here, the signatures of the rare classes are very similar to those of the larger classes, unlike the contrasting signature of the class U in the previous case. Trying to find these rare classes among the 20 classes using the same 10 x 10 SOM is a larger data mining challenge than finding the U class in the 5-class image. Correlation coefficients between the different dimensions range from $0.0081(\rho_{v_5v_4})$ to $0.5641(\rho_{v_4v_2})$, so this too is a "forbidden" case for BDH according to the available theory. Clustering this data set using KSOM is depicted in Figure 7, left. The rare classes are detectible but each is poorly represented, only by a single PE. Also, because of the tight quarters (same SOM size for more classes) the separating fences (the contrasts between the PE weights) are less pronounced, which makes the small classes less discernible. Figure 7, right, shows that BDH with $\alpha_{desired} = -0.8$, magnifies the rare classes in the map: the 1-pixel class R is now represented by 4 PEs in contrast to 1 PE in the map formed by KSOM, and class Q is represented by 7 PEs. The detectability of the rare classes has been increased by magnified areal representation as well as by higher contrast in the weight differences.

The aim of negative magnification is to enhance the SOM areal representation of low-density data regions by forcing the allocation of more prototypes to those regions. This may improve the detectability of unknown rare classes. Naturally, the representation of the non-rare classes in such a map will be somewhat repressed as is apparent from Figure 7, bottom. The map obtained with $\alpha_{achieved} < 0$ and the one obtained with $\alpha_{achieved} \approx 1$ together provide a more complete picture of the structure of a data set.

In order to further chart the behavior of the BDH, we now present additional controlled experiments with synthetic as well as real data sets. Beside $\alpha < 0$, of special interest is the case of $\alpha = 1$ because it effects maximum entropy quantization, and thus helps faithful mapping of the input data structure.

B. Inducing $\alpha = 1$ magnification on 6-dimensional synthetic data

After demonstrating that negative magnification can be achieved on higher-dimensional data, now we show that $\alpha = 1$ can also be achieved by BDH, fairly accurately, on similar data. $\alpha = 1$ is a special case in that the heuristic Conscience algorithm by DeSieno [6] is believed to achieve maximum entropy quantization, therefore, if we verify that, we can compare the properties of the SOM obtained by BDH magnification with an SOM obtained by Conscience, to determine if the BDH realized $\alpha = 1$. In addition, we can compare the entropies of the two maps.

The simulations presented in this section serve two purposes. One is to demonstrate that the Conscience algorithm indeed produces *pdf* matching, and the second is to show that

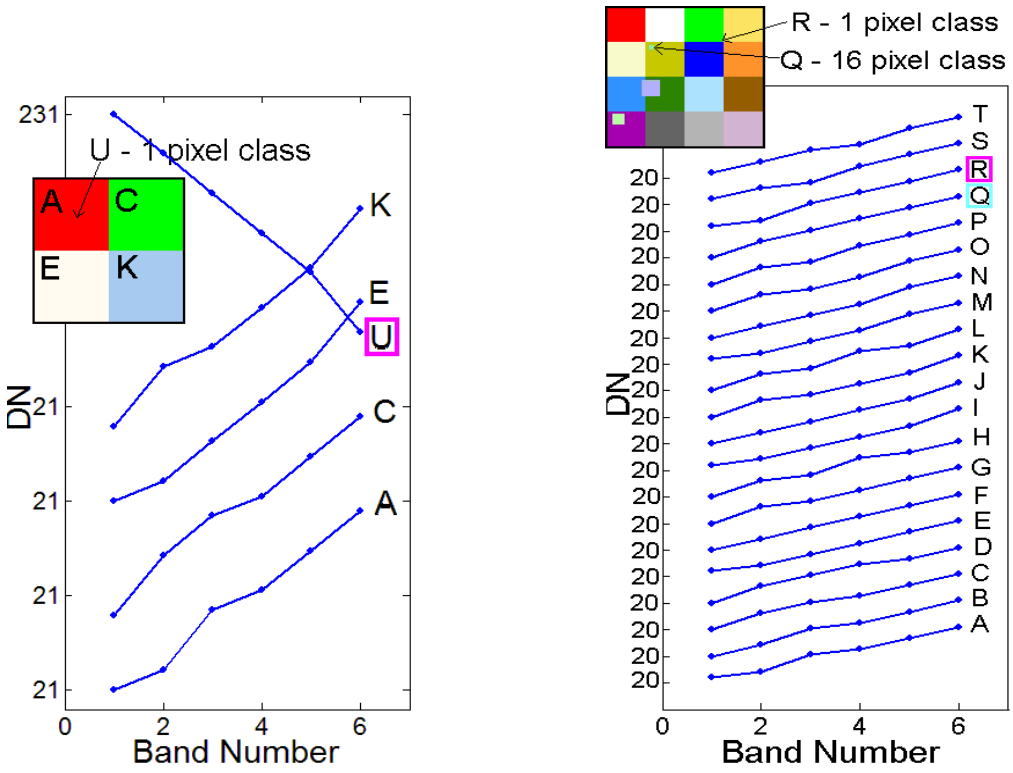

Fig. 5. 6-dimensional synthetic image data sets: the known class structure and spectral signatures of the classes. Each spectral signature is displayed with an offset for clear viewing. The Data Numbers (DN) on the y axis indicate the values of the first band of each respective spectrum. The spatial area of both images is 128 x 128 pixels. **Left:** 5-class image data set with four large square classes (A, C, E, K, with 4095, 4096, 4096 and 4096 pixels, respectively), and one 1-pixel class, U. The spectral signature of class U is dramatically different from the rest. **Right:** 20-class image data set, with 16 nearly equal area (32 x 32 pixels) classes A - P, and four additional very small classes: R, Q, S, T contain 1, 16, 64 and 128 pixels, respectively. Classes Q - T take away the corresponding number of pixels from those of the A - P classes in which they are embedded.

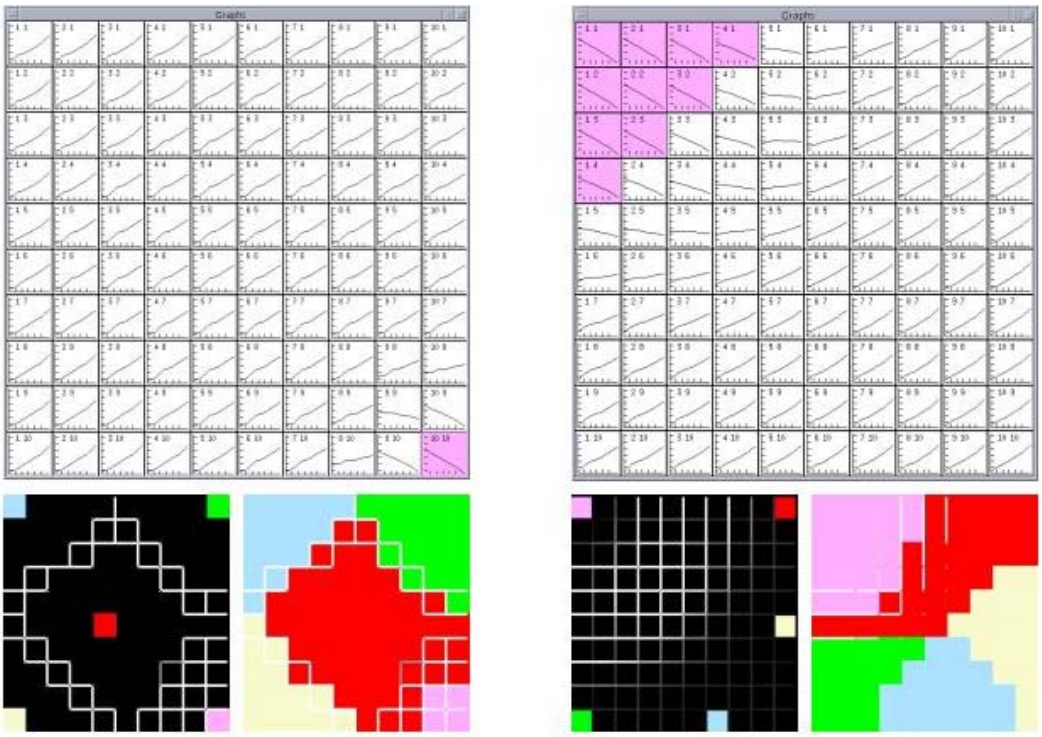

Fig. 6. Magnification results with the 5-class data set. **Left panel:** Using KSOM. **Top:** Weight vectors of the 10×10 SOM. Only 1 PE represents the rare class U in the lower right corner. **Bottom Left:** Clusters identified in the map by visualizing the weight distances of adjacent PEs (the darker the fence between two PEs, the smaller the difference between the corresponding weights). **Bottom Right:** This figure shows which class each weight vector is closest to, which complements the information on the left. **Right panel:** Using BDH with $\alpha = -0.8$. **Top:** Weight vectors in the 10×10 SOM. The rare class U is now represented by 10 PEs! **Bottom Left:** Clusters identified in the SOM. **Bottom Right:** The map of nearest class for each weight vector.

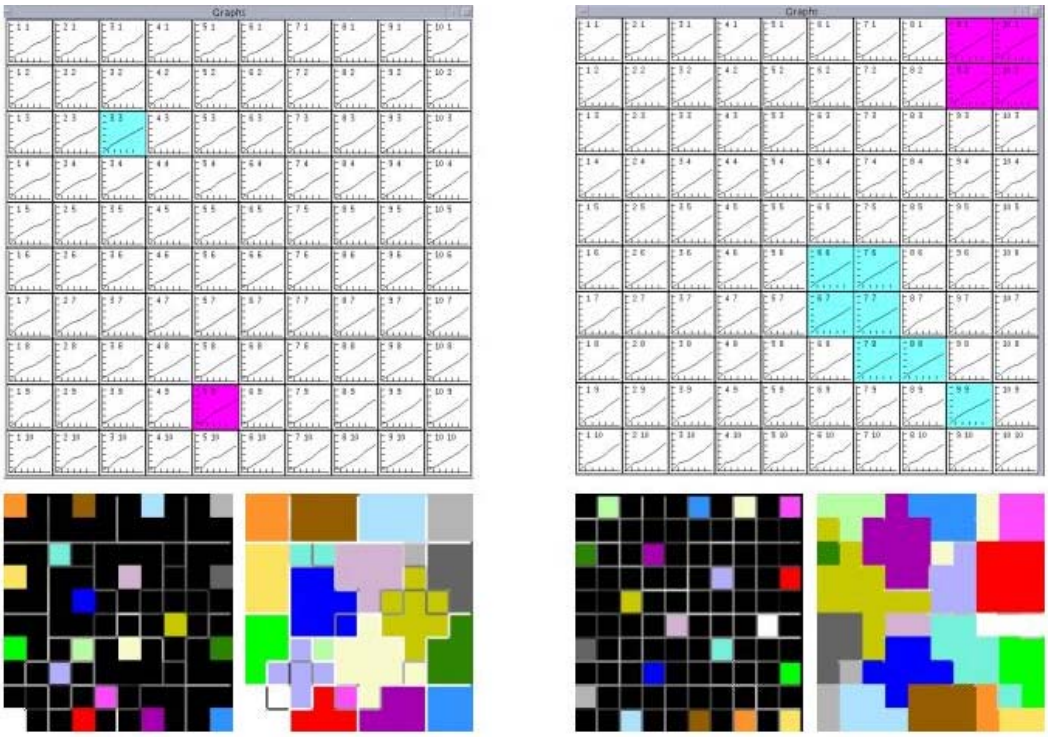

Fig. 7. Magnification study with the 20-class image. **Left panel:** Using KSOM. **Top:** Weight vectors in the 10×10 SOM. Only 1 PE is allocated for each of the rare classes R and Q. **Bottom Left:** Clusters identified in the SOM. Fence intensity from black to white is proportional to the distance between the corresponding weights. **Bottom Right:** This figure shows which class each weight vector is closest to, which complements the information on the left. **Right panel:** Using BDH with $\alpha_{desired} = -0.8$. **Top:** Weight vectors in the 10×10 SOM. Now 4 and 7 PEs represent the rare classes R and Q, respectively. **Bottom Left:** Clusters identified in the SOM. **Bottom Right:** Map of closest class for each weight vector.

by forcing $\alpha = 1$ the BDH can accurately model the *pdf*, at least on the level of the cluster structure of the data. We do this by a direct comparison of the clusters detected by the SOMs with the known labels of data points (the known clusters) in the artificially generated data. In many data mining pursuits, among them remote sensing and medical applications, the main objective is to find interesting, relevant groupings in the data, but not necessarily a finer approximation of the *pdf*. Such challenges can be met without a very precise evaluation of α.

The data set we use for this purpose is an 8-class synthetic image consisting of 6 image bands, similar to the data described in Figure 5 except that here 8 spectral types are distributed over subareas of the 128 x 128 pixel image in the following manner: Classes A and B each cover 4096 pixels, classes C and O are each 2048 pixels, and classes D, H, I, M have 1024 pixels. Gaussian noise, about 10% on average, was added to create more realistic variations within the spectral classes. The Conscience algorithm is expected to map each of these classes onto areas in the SOM that are proportional in size to the areas of the classes, namely classes D, H, I, and M should each occupy half as many Processing Elements than either of class C or O, and A and B both should be represented by twice as many PEs as C or O, and by four times as many PEs as any of D, H, I, M. Figure 8 shows that indeed, this is the case within the accuracy allowed by the size of the 15 x 15 square SOM grid, by integer arithmetic, and taking into account the empty PEs that form dividing gaps between clusters. The clusters are captured using a somewhat modified version of the U-matrix [18], in that we compute and visualize the distances to the neighbor weights on either side of a PE separately, and do this also for the diagonal neighbors (we use an 8-neighbor neighborhood). The intensity of the "fences" between PEs is proportional to the distance between weights, on the black to white gray scale. White is large difference, black is great similarity. Clusters are clearly outlined by the white fences. Fences within the clusters are uniformly very low (virtually 0). Out of 225 SOM PEs, classes A and B cover 48 and 49 PEs, C and O cover 25 and 21 PEs, and the smallest four classes occupy 13, 9, 10, and 9 PEs, respectively. 41 PEs belong to inter-cluster gaps. The largest deviation from the expected values occurs in the smallest classes, which is understandable considering that just one additional PE in each of H and M, taken away from D would even out the areas to 10 – 11 PEs each. The SOM was run for 2 million steps to ensure convergence, but the cluster structure in Figure 8 was already formed after 2-300,000 steps. It is noteworthy that the precision of the *pdf* estimation is better than the level of the cluster structure: the intensity of the red color in each PE cell on the left of Figure 8 is proportional to the relative winning frequency of the respective PE, and as is evident the red shades are fairly uniformly distributed within clusters.

Since a maximum Shannon entropy (equiprobabilistic) mapping provides the most faithful match of the *pdf* by the given number of prototypes, the entropy of the SOM is a good indicator of the quality of the learned quantization. The

Shannon entropy of the SOM is given by

$$I = \sum_{i=1}^{N} P(j) log_2 P(j) \quad (10)$$

where j is the PE index in the SOM ($j = 1, ..., N$), $P(j) = \frac{I_j}{I_N}$ is the probability that PE j wins an input sample, where I_j is the number of inputs samples mapped to (won by) PE j and I_N is the total number of input samples. The higher the value of the map entropy, the closer the mapping is to equiprobabilistic. In an equiprobabilistic map $P(j) = \frac{1}{N} \quad \forall j$, and in that case I would have a maximum value of $log_2 N$. We compute a normalized entropy by dividing I by the theoretical maximum, $log_2 N$ (thus the maximum possible value of the entropy of any map is 1), so that entropies can be compared across SOMs of different sizes. The (normalized) entropy of the Conscience SOM in Figure 8, is 0.998 for the active PEs, and 0.962 for all PEs.

After this verification of the expected performance of the Conscience algorithm we can evaluate the magnification performance of the BDH on the same data. The pairwise correlations of the various dimensions of this data set are typically strong, (most are between 0.3 - 0.8, and only two are less than 0.05). Based on the 2-d experiments summarized in Figure 4, and lacking any guidance from theory or other works, we hypothesize that a similar trend (as in Figure 4) may apply for higher-dimensional forbidden data, *i.e.,* in order to produce a certain $alpha_{achieved}$ value one needs to request an $alpha_{desired}$ value that is offset according to some observed function. Since this work is the first one to chart such functions the most reasonable assumption we can make is the functional relationship we charted for 2-dimensional data with strong dimensional correlations, and correct our assumption according to the outcome of the simulation, if needed. By extrapolation of Figure 4, we assume that in order to achieve $\alpha_{achieved} = 1$ the BDH needs to be run with a choice of $\alpha_{desired} = 0.7$. The SOM, formed by BDH magnification forcing of $\alpha_{achieved} = 1$ in this way (Figure 9), shows that it has a very similar area distribution over the 8 classes as the Conscience SOM in Figure 8 (again, with similar accuracy considerations). This indicates that the BDH fairly closely achieved the desired maximum entropy mapping on this "forbidden" data set. The entropy of this map (relative to the theoretical maximum) is 0.96 for the active PEs, 0.914 for all PEs. We note that the entropy of the BDH SOM is lower than that of the Conscience SOM. This is also obvious from the density of the cells that are close to the cluster boundaries in Figure 9. By preliminary observation, the entropy of the BDH SOM very slowly increases over a very long run, however, it does not reach the entropy of the Conscience SOM in approximately 7 million steps. We think that this discrepancy could be due to other circumstances that the BDH may be more sensitive to, such as the combination of the number of clusters and the size of the SOM lattice, or the learning schedule including the decay rate of the neighborhood. This needs to be evaluated in subsequent work.

One difficulty with the BDH is the estimation of the data and weight density at the locations of the weights (equation (4)) [19]. The estimation (equation (5)) uses a rough approximation of the volume of the receptive field of $\mathbf{w_s}$, $||\mathbf{v} - \mathbf{w_s}||^d$ where d is the *effective dimensionality* of the receptive field. This dimensionality may vary from weight to weight and is usually unknown, therefore any global value of d is a crude estimate. For that reason we can expect the BDH to work well with a range of d values (within $[1, n]$). Therefore we chose $d = 2$, which is the squared Euclidean distance and thus has the advantage of saving on computational cost. We add that the relative insensitivity of the BDH to the value of d agrees with our experiences with a limited number of trials.

C. Finding rare clusters in a real spectral image

As we mentioned in the introduction, magnification control can lead to the detection of rare clusters. In particular, for inverted (negative) magnification prototypes are preferentially placed in low density areas of the data space, resulting in more accurate description of those low probability regions. Low density regions may contain meaningful, separate clusters, which are not detectable if there are not enough representing prototypes but may become "visible" through better representation using negative magnification. We will demonstrate this effect with a real world data example, which is an urban remote sensing spectral image of Ocean City, Maryland. We use a 512 x 512 pixel, 8-band subset of the Ocean City image to study the effect of forced negative magnification. This data set also has high pairwise correlations, the magnitudes of which are mostly between 0.5 and 0.95. We cannot compute the value of $\alpha_{achieved}$, but we can compare the appearance 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 than in the Conscience SOM. In addition, we look for previously unidentified clusters. An earlier supervised classification that was independently verified against ground truth provides the knowledge of existing clusters in the data. Figure 10 demonstrates the discovery of one new - very small - cluster. It also shows another small cluster (pale aqua, class V) that was known at the time of the earlier supervised classification, but was more definitely outlined by BDH clustering. Figure 11 compares the two SOMs. Shown on the left is the 40 x 40 SOM formed by BDH learning with $\alpha_{desired} = -0.8$, using only the upper right quadrant of the image (framed in Figure 10), *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 distinctively different from all other clusters as seen in Figure 12. 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 in Figure 10. The 40 x 40 SOM produced by Conscience learning, using the *entire* image, is in the middle. The greenish-yellow cluster was hard to see in this map, and was only "discovered" 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 contoured by better

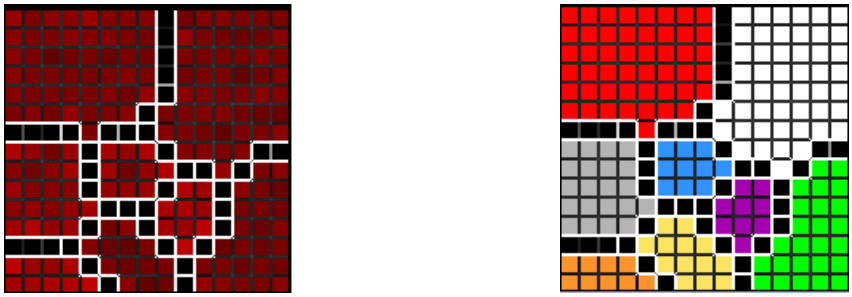

Fig. 8. The 6-dimensional 8-class data set as represented by an SOM that learned by the Conscience algorithm. **Left:** The cluster boundaries, visualized as the distance of weights of adjacent PEs, similarly to that of the U-matrix [18] except we compute and display the distances to all 8 neighbors separately, by the fences on the edges and corners of the PE grid cells. White is high fence (large dissimilarity), black is low fence (great similarity). The SOM has a 15 x 15 square grid. Each grid cell is shaded by an intensity of red proportional to the number of data points mapped to the PE in that grid cell. Black grid cells between the strong fences indicate that the receptive fields of the corresponding PEs are empty. The fairly uniform density over the non-empty PEs indicate a good approximation of maximum entropy mapping. The entropy of this map (relative to the theoretical maximum) is 0.998 for the active PEs, 0.962 for all PEs. **Right:** The known class labels superimposed over the PE grid cells. Both representations show that the PEs (and SOM weights) are divided among the classes in proportion to the sizes of the classes: A, B (red and white) contain 4096 data points each, C, O (green and grey) 2048, and D, H, I, and M have 1024 points. The corresponding number of designated PEs are A:48, B:49, C:25, O:21, D:13, H:9, I:10, M:9. The deviations from the exact 4:2:1 proportions can be due to the small size of the SOM, integer arithmetic, and the formation of inter-cluster gaps.

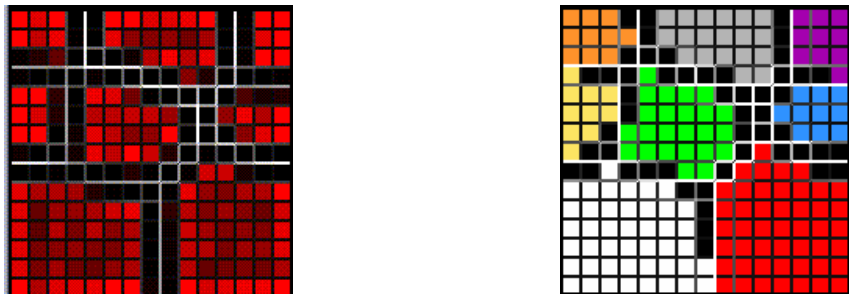

Fig. 9. The 8 known classes of the same 6-dimensional patterns as in Figure 8, as represented by an SOM that learned via the BDH magnification control, forcing $\alpha = 1$. Since this is a "forbidden" data set with strong inter-dimensional correlations, inducing $\alpha_{desired} = 0.7$ effectively produced $\alpha_{achieved} = 1$, as suggested by the experiments in Figure 4. **Left:** The cluster boundaries and data density, visualized the same way as in Figure 8. The entropy of this map (relative to the theoretical maximum) is 0.96 for the active PEs, 0.914 for all PEs. Note that many of the dark grid cells on either sides of single line fences have data points mapped to them albeit few. **Right:** The known class labels superimposed over the PE grid cells. Both representations show that the PEs (prototype vectors) are divided among the classes in proportion to the sizes of the classes, similarly to the Conscience algorithm results in Figure 8: A, B (red and white) contain 4096 data points each, C, O (green and grey) 2048, and D, H, I, and M have 1024 points. The corresponding number of designated PEs are A:47, B:44, C:24, O:19, D:10, H:10, I:10, M:10. The deviations from the exact 4:2:1 proportions can be due to the small size of the SOM, integer arithmetic, and the formation of inter-cluster gaps.

developed "fences". Similarly, the previously known small V class is represented by 4 PEs in the Conscience SOM vs 6 PEs in the BDH SOM, even though the Conscience SOM was learned with 4 times as many data points, including more occurrences of the V class in the large image outside the upper quadrangle, and as seen from the image on the right in Figure 11. The previously knwon white class 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 (most notably the long vertical rectangle in the lower right image corner). These observations clearly indicate that, compared to Conscience the SOM, the BDH preformed negative magnification.

IV. Discussion of problems, and conclusions

We mention here some issues related to this work but not discussed above.

All of our evaluations assume topologically correct mapping by the respective SOMs (that the learning does not result in "twisted" maps.) Ensuring and verification of this is not easy, especially for high-dimensional, complex data. Although supporting theories exist (Bauer and Pawelzik [20], Villmann *et al.* [21]), those, similarly to the BDH itself, need systematic evaluation for higher-dimensional data. In these applications we used data of which we knew the cluster structure from independent investigations, thus we could verify the validity of the clusters identified by the SOM.

In our experience, finding "good" learning parameters is more difficult for the BDH algorithm than for KSOM or for the Conscience algorithm. One especially important question seems to be the value of negative α: can one "over-distort" the rest of the map while discovering rare clusters, by using a too large negative value? In several cases of real data we saw that while extremely rare classes were beautifully magnified and further separated into meaningful subclusters such as in discovering rare Martian mineral types [22], the rest of the data were mapped onto thin filament-like groups of PEs leaving most of the SOM grid empty.

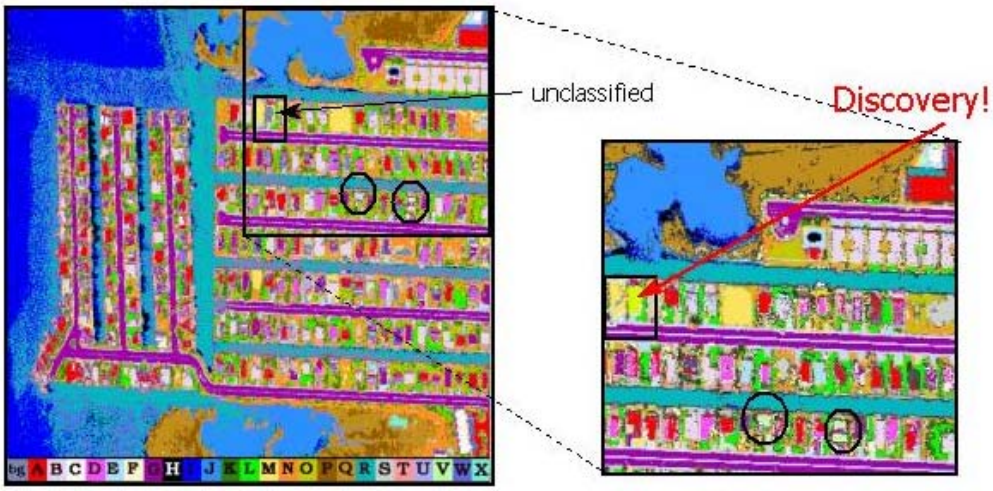

Fig. 10. Comparison of supervised classification and BDH clustering with $\alpha < 0$, on an 8-band spectral image of Ocean City, Maryland. **Left:** An earlier supervised classification that satisfactorily mapped 24 known cover types of interest, based on verification against ground truth. Shown centered 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, to the right of a yellow rectangular patch. He pale aqua spots in the black ovals belong to the rare class V, known at the time of the supervised classification. **Right:** SOM clustering using BDH magnification control with $\alpha_{desired} = -0.8$ 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 color (greenish-yellow) that is different from all previous class colors: the spectral signature of this area is distinct. We discovered a new class. Moreover, this cluster only occurs at this location, and nowhere else: we discovered a small rare class! Figure 11 shows the SOM view of this discovery. Note that the formerly known rare class V (pale aqua spots in black ovals) became better defined. Class spectral signatures are plotted in Figure 12.

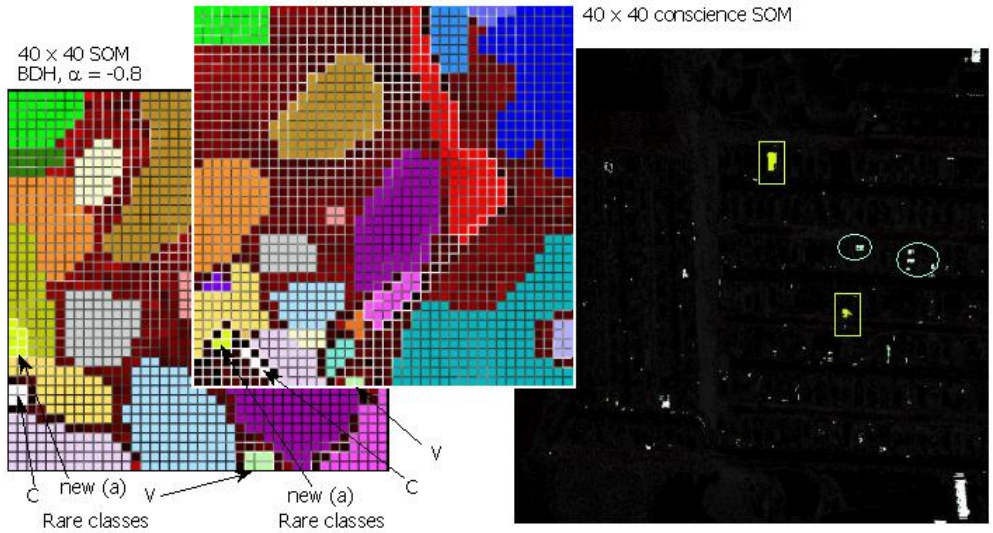

Fig. 11. Comparison of SOMs developed by BDH versus Conscience learning. **Left:** The SOM learned by BDH, $\alpha_{desired} = -0.8$, using the upper right quadrant of the 512 x 512 pixel Ocean City image, shown framed in Figure 10. **Middle:** The SOM learned by the Conscience algorithm ($\approx \alpha = 1$), using the entire Ocean City image. **Right:** The rare classes in the image. The newly fond cluster (indicated in the rectangles) only occurs in one other spot in the entire image, outside of the BDH-discovered location within the upper right quadrangle. It is apparent, as explained in the text, that the rare clusters are magnified in the BDH SOM in comparison to the Conscience SOM.

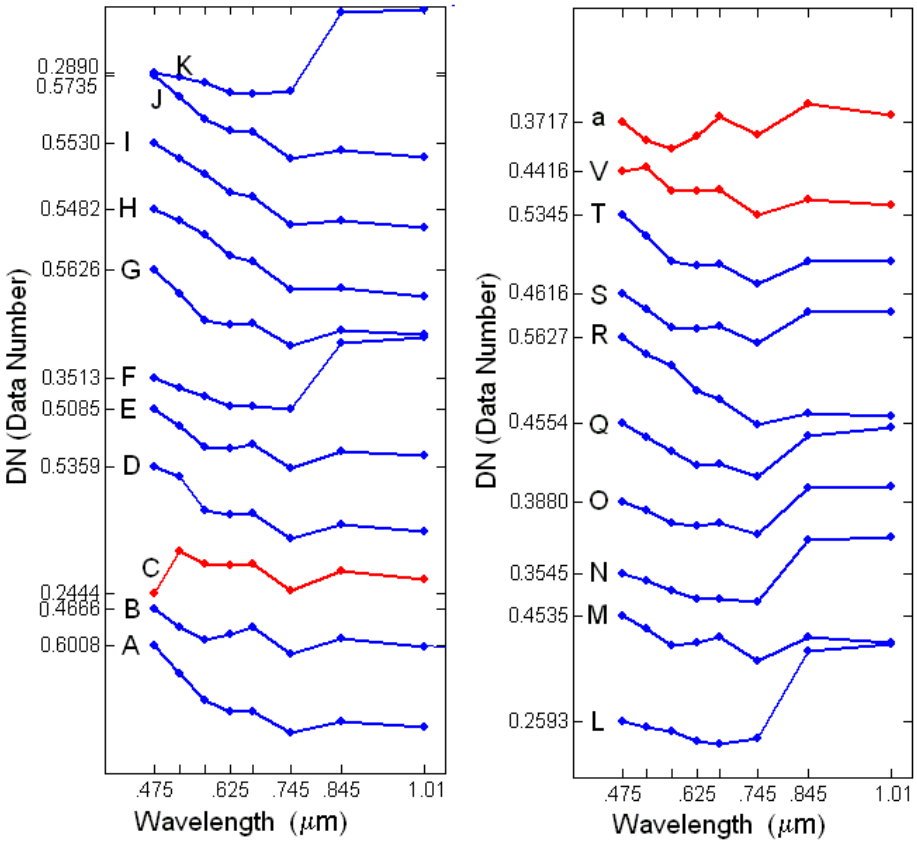

Fig. 12. Mean spectral signatures of the clusters identified in the Ocean City spectral image. Spectra are offset for clarity. The DN value on the vertical axis indicates the value in the first spectral band. The newly discovered cluster (greenish-yellow in Figure 10) is labeled "a" on top of the right panel. The spectral dissimilarity of this cluster with the rest is obvious. The other rare classes discussed in the text are class C and V.

A related item is that it is hard to estimate what $\alpha_{desired}$ value should be used as input to the BDH, in order to achieve a specific α value. The experiments presented in Figure 4 suggest a trend but we caution against straight extension of those values to higher-dimensional data. Some preliminary experiments with the 8-class data seem to indicate that $\alpha = 1$ may successfully be induced within a range of $\alpha_{desired}$. We found that to be the case for $0.6 \leq \alpha_{desired} \leq 0.9$, *with this particular data set*. It may be worth further simulations to get a firmer grasp on stability regions (that we suspect may exist), especially for negative values. The current lack of capability to evaluate $\alpha_{achieved}$ from the SOM of high-dimensional data makes this problem even harder.

The above considerations suggest that magnification control may influence other properties of the SOM, namely, the ability of topographic mapping. Thus, if one is interested in optimizing both magnification and topographic mapping, one has to balance between these aspects and to prove the topography by appropriate tools. Alternatively, the magnification scheme can be integrated directly into the Growing Self-Organizing Map (GSOM) approach [23].

The estimation of the data and weight density at the locations of the weights (eq. 5) is problematic in the BDH algorithm, as we mentioned in Section III.B. A more accurate assessment may be possible by the estimation of the volume of the Voronoi cells, which may be done through the size of the respective receptive fields (that the algorithm can record continuously) and may result in marked improvement of the algorithm. This is a follow-up task for us, worth pursuing in our opinion, especially since the BDH approach seems to be the most convenient method to control the magnification (see [12]).

Of course, the most interesting and outstanding issue is the theoretical justification of magnification control for higher-dimensional cases.

Kohonen proposed a theoretical approach for higher dimensions by modeling the local receptive field densities with simple hyperspheres of constant data density, which enables local factorization [19]. However, this assumption is generally not true for real data. In addition, Kohonen's considerations require that the SOM dimensionality match that of the data. A modification of the SOM winner determination by Heskes [24] leads to an energy function for the learning of SOMs, valid also for higher dimensions, but derivation of magnification properties seems to be very difficult because in the Heskes scheme the neighborhood function influences the winner selection. If we were to follow the derivation of Ritter and Schulten [4] a recursive equilibrium equation would result for which no solution exists at present. Another

extension of SOMs is the Neural Gas paradigm whose learning dynamics also has an energy function [25]. For the neural Gas, analytical proofs exist for several different magnification control paradigms including the BDH learning. However, since the energy function is obtained by the modification of the neighborhood function such that it is evaluated in the data space, the neighborhood structure among PEs is lost.

The difficulties illuminated by these investigations further motivate the numerical simulations we described in this work.

In conclusion, we presented systematic experiments with the map magnification control by Bauer, Der, and Hermann (BDH) [1], on data for which the BDH scheme is not supported by existing theory. Based on our observations of the systematic BDH behavior on 2-dimensional non-separating "forbidden" data, we were able to induce maximum entropy ($\alpha = 1$) mapping and negative magnification on 6-dimensional synthetic data. We also showed that negative magnification worked on 8-dimensional real image data and helped enhance the areal representation, and thereby the discovery of rare clusters. While the range of our studies is too limited to draw definite conclusions, the simulations indicate *consistent* behavior of the BDH on some set of "forbidden" data. This encourages further simulations to investigate the predictability of the BDH for potential analyses of complex, high-dimensional data. That, in turn, perhaps can inspire more theoretical studies and algorithm development.

APPENDIX: A HISTOGRAM BASED METHOD FOR EVALUATION OF α

The following power law relates the density of weights in the input space $Q(\mathbf{w})$ to the *pdf* $P(\mathbf{w})$ of the input samples,

$$Q(\mathbf{w}) = P(\mathbf{w})^{\alpha} \times const \tag{11}$$

where $Q(\mathbf{w})$, the density of weights in the input space, is the number of reference vectors in a small volume $d\mathbf{w}$ of the input space. If both sides of Equation (11) are divided by N_W, the total number of weight vectors, on the left we get $Q(\mathbf{w})/N_W$, the *pdf* of the weight vectors and on the right, $const/N_W$ can be absorbed into a new constant $constant = const/N_W$. From now on we will use the same power law as in Equation (11), but with the understanding that $Q(\mathbf{w})$ now denotes the *pdf* of the weight vectors.

From Equation (11), it is clear that finding the value of α ($\alpha_{achieved}$) requires the estimation of the two *pdf*s. The following histogram based method can be used to estimate the *pdf*s and evaluate α.

Let p_i and q_i be samples of the two densities $P(\mathbf{w})$ and $Q(\mathbf{w})$ respectively. p_i and q_i can be obtained from the input samples and weights by partitioning the input space into N_B bins and constructing frequency histograms in the following way:

$$q_i = \left(\frac{n_{W_i}}{N_W}\right)/V_i, \quad i = 1, 2, \ldots, N_B \tag{12}$$

where n_{W_i} is the number of weights in the the i^{th} bin, N_W is the total number of weights and V_i is the volume of the i^{th} bin.

$$p_i = \left(\frac{n_{I_i}}{N_I}\right)/V_i, \quad i = 1, 2, \ldots, N_B \tag{13}$$

where n_{I_i} is the number of input samples in the the i^{th} bin, N_I is the total number of input samples and V_i is the volume of the i^{th} bin.

The value of α that best satisfies

$$q_i = constant \times p_i^{\alpha}, \quad i = 1, 2, \ldots, N_B \tag{14}$$

in other terms, one that minimizes the error measure,

$$E(\alpha) = \sum_{i=1}^{N_B} (q_i - constant * p_i^{\alpha})^2 \tag{15}$$

is the value of $\alpha_{achieved}$ by the SOM. $\alpha_{achieved}$ can be found by varying α in a range around the value of $\alpha_{desired}$, determine the corresponding value of the error $E(\alpha)$ and then choose the $\alpha_{achieved}$ as that value of α that minimizes $E(\alpha)$.

The $constant$ for a particular value of α can be determined by noting that $\sum_{i=1}^{N_B} q_i * (V_i) = 1$ (where $q_i * (V_i)$ is the probability that a weight lies in the i^{th} bin, summing this over all bins yields 1). So multiplying both sides of Equation (14) by V_i and summing over N_B bins gives

$$constant = \frac{1}{\sum_{i=1}^{N_B} p_i^{\alpha}} \tag{16}$$

This histogram based method for determining the value of $\alpha_{achieved}$ seems simple but there are difficulties in extending it to higher dimensions. The *bin* in a d-dimensional space is a hypercube. In order to have an accurate estimation of the *pdf* and hence that of $\alpha_{achieved}$, it is important to determine the correct size of the bin. As it turns out, this is not an easy task! The number of input samples required for estimating either of the *pdf*s for high-dimensional data is prohibitive. As a consequence of this, the evaluation of $\alpha_{achieved}$ in a general high-dimensional case is a difficult problem. For reasons of space constraints, we refer the reader to [26] for discussion of bin size selection and other issues involved in *pdf* estimation.

ACKNOWLEDGEMENTS

This work was partially supported by grants NAG9-10432 and NNG05GA94G from the Applied Information Research Program of NASA, Science Mission Directorate. Many thanks to Dr. Bea Csathó, Byrd Polar Institute, Ohio State University, for the Ocean City image and ground truth. Philip Tracadas, Rice University, contributed greatly with software development for the HyperEye © environment (URL http://www.ece.rice.edu/HYPEREYE), in which the presented higher-dimensional simulations were run.

REFERENCES

[1] 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.

[2] P.L. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. on Information Theory*, vol. 28, no. 2, pp. 139–149, March 1982.

[3] T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, New York, 1988.

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

[5] D.R. Dersch and P. Tavan, "Asymptotic level density in topological feature maps," *IEEE Trans. on Neural Networks*, vol. 6, no. 1, pp. 230–236, January 1995.

[6] D. DeSieno, "Adding a conscience to competitive learning," in *Proc. Int'l Conference on Neural Networks (ICNN), July 1988*, New York, 1988, vol. I, pp. I–117–124.
[7] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin Heidelberg New York, 1997.
[8] A. Jain and E. Merényi, "Forbidden magnification? I.," in *Proc. of European Symposium on Artificial Neural Networks (ESANN'04)*, Bruges, Belgium, April 28-30, 2004, pp. 51–56, D facto publications.
[9] E. Merényi and A. Jain, "Forbidden magnification? II.," in *Proc. of European Symposium on Artificial Neural Networks (ESANN'04)*, Bruges, Belgium, April 28-30, 2004, pp. 57–62, D facto publications.
[10] H. Ritter, "Asymptotic level density for a class of vector quantization processes," *IEEE Trans. on Neural Networks*, vol. 2, pp. 173–175, 1991.
[11] A. Jain, *Issues Related to Data Mining with Self-Organizing Maps*, Rice University, 2004, MS Thesis.
[12] Th. Villmann and J.C. Claussen, "Magnification control in self-organizing maps and neural gas," *Neural Computation*, vol. 18, pp. 446–469, 2006.
[13] S. C. Ahalt, A. K. Krishnamurty, P. Chen, and D. E. Melton, "Competitive learning algorithms for vector quentization," *Neural Networks*, vol. 3, no. 3, pp. 277–290, 1990.
[14] T. Villmann, "Controlling strategies for the magnification factor in the neural gas network," *Neural Network World*, vol. 10, no. 4, pp. 739–750, 2000.
[15] M. M. Van Hulle, *Faithful Representations and Topographic Maps*, Wiley Series and Adaptive Learning Systems for Signal Processing, Communications, and Control. Wiley & Sons, New York, 2000.
[16] J. Lampinen and T. Kostiainen, "Generative probability density model in the self-organizing map," in *Self-Organizing Neural Networks*, U. Seiffert and L.C. Jain, Eds., Studies in Fuzziness and Soft Computing, pp. 75–92. Physica-Verlag, Heidelberg, New York, 2001.
[17] T. Villmann, E. Merényi, and B. Hammer, "Neural maps in remote sensing image analysis," *Neural Networks*, vol. 16, pp. 389–403, 2003.
[18] 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.
[19] T. Kohonen, "Comparison of SOM point densities based on different criteria," *Neural Computation*, vol. 11, pp. 20813/42095, 1999.
[20] 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.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] D.W. Scott, *Multivariate Density Estimation-Theory,Practice and Visualization*, A Wiley-Interscience Publication, 1992.