

Forbidden Magnification? II.

Erzsébet Merényi and Abha Jain *

Rice University, Electrical and Computer Engineering Department, MS-380

6100 Main Street, Houston TX 77005, U.S.A.

e-mail: {erzsebet,abha}@rice.edu

Abstract. The twin of this paper, “Forbidden Magnification? I.” [1], presents systematic SOM simulations with the explicit magnification control scheme of Bauer, Der, and Herrmann [2] on data for which the theory does not guarantee success, namely data that are n -D, $n > 2$ and/or data whose components in the different dimensions are not statistically independent. For the unsupported $n = 2$ cases that we investigated the simulations show that even though the magnification exponent $\alpha_{achieved}$ achieved by magnification control is not the same as the intended $\alpha_{intended}$, the direction and sign of $\alpha_{achieved}$ systematically follows $\alpha_{intended}$ with a more or less constant offset. We experimentally showed that for simple synthetic higher dimensional data negative magnification has the desired effect of improving the detectability of rare classes. In this paper we study further theoretically unsupported cases, including experiments with real data.

1 Known limits of SOM magnification control

Controlling the magnification of Self-Organizing Neural Maps (*i.e.* the functional relationship between the *pdf* of the input data and the density of the SOM weights in the input space) is an extremely attractive possibility because various values of the magnification exponent, denoted by α in this paper, effect desirable quantization properties. For example, $\alpha = 1$ realizes maximum entropy quantization, $\alpha = 1/3$ and $\alpha = 1/2$ force minimum distortion quantization for 1- and 2-D data, respectively. As is known, the basic Kohonen SOM's inherent property is a map magnification of $\alpha = 2/3$. The twin of this paper, [1], and references therein give more details on magnification. The algorithm by Bauer, Der, and Herrmann [2] (referred to as BDH from now on) provided a principled approach to obtaining a desired magnification exponent for 1-D data and for 2-D data whose components are statistically independent. Most real data, of course, do not obey the above conditions, yet it is real data scenarios that would benefit the most from explicit magnification control. After careful verification of known magnification properties on “allowed” data we examined

Request color reprint by email.

*Authors are partially supported by the Applied Information Systems Research Program of NASA, Office of Space Science, NAG9-10432.

the effects of the BDH method on simple “forbidden data” in [1]. Within the range that we observed, the magnification exponent, $\alpha_{achieved}$, obtained by BDH, shows a systematic difference from the intended magnification exponent $\alpha_{intended}$, which suggests that the BDH may be justified for a larger range of data than supported by available theories. In order to chart the behavior of the BDH, in this work we present additional controlled experiments with synthetic as well as real data sets. We are especially interested in the cases of $\alpha = 1$ and $\alpha < 0$. $\alpha = 1$ effects maximum entropy quantization, and thus helps faithful mapping of the input data structure. $\alpha < 0$ enlarges the representation of small classes in the SOM due to which detection of those rare classes becomes easier, discoveries of important, subtle anomalies may be possible that otherwise could remain hidden. Both of these properties can be particularly useful for complex, high-D data if systematic experiments indicate a predictable behavior of the BDH for such “forbidden data”.

2 Inducing $\alpha = 1$ magnification on 6-D data

[1] demonstrated the desired effect of negative magnification for a 20-class 6-D data set. Even though we could not compute the exact value of $\alpha_{achieved}$ it is of great value to know that the forcing of a negative $\alpha_{intended}$ did make rare clusters significantly more visible in the SOM. Now we show that $\alpha = 1$ can be achieved by BDH, fairly accurately, on similar data. $\alpha = 1$ is a special case in that the heuristic “Conscience” algorithm by DeSieno [3] is believed to achieve maximum entropy quantization, therefore we can compare the properties of the SOM obtained by Conscience, and the SOM obtained by BDH magnification. The data set we use for this purpose is an 8-class synthetic image consisting of 6 image bands, similar to Data Set II described in [1] except that here 8 of the 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 (PEs) 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 1 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. 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 1 was already formed after 2-300K steps.

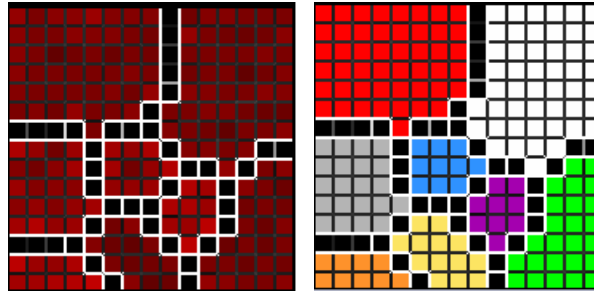


Figure 1: The 8 known classes of 6-D patterns, as represented by an SOM that learned by the Conscience algorithm. **Left:** The cluster boundaries, visualized as the distance of weights of adjacent PEs. White is high fence (large dissimilarity), black is low fence (great similarity). The SOM has a 15 x 15 rectangular 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. **Right:** The known class labels superimposed over the PE grid cells. Both representations show that the PEs (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. Original figure is in color. Download paper from <http://www.ece.rice.edu/~erzsebet/papers/esann04-2.pdf>.

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 between 0.3 - 0.8, and only two less than 0.05). Therefore, based on Figure 4 of [1] we hypothesize that in order to produce $\alpha_{achieved} = 1$ the BDH needs to be run with a choice of $\alpha_{intended} = 0.7$. The SOM, formed by BDH magnification forcing of $\alpha_{achieved} = 1$ in this way (Figure 2), shows that it has a very similar area distribution over the 8 classes as the Conscience SOM in Figure 1 (again, with similar accuracy considerations). This indicates that the BDH fairly closely achieved the desired maximum entropy mapping on this “forbidden” data set. One problem with the BDH is the need to estimate the effective dimensionality of the data. Here, we used $d=2$.

3 Finding rare clusters in a real spectral image

An urban remote sensing spectral image of Ocean City, Maryland, is used for our first demonstration with *real* 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 -D vector, called the spectrum, which carries compositional information of the material in the

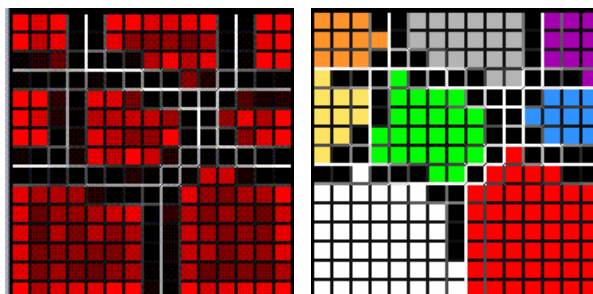


Figure 2: The 8 known classes of the same 6-D patterns as in Figure 1, as represented by an SOM that learned via the BDH magnification control, forcing $\alpha = 1$. Since this is a “forbidden” $n > 2$ dimensional data set with strong inter-dimensional correlations, inducing $\alpha = 0.7$ effectively produced $\alpha_{achieved} = 1$, as suggested by the experiments in Figure 4 of [1]. **Left:** The cluster boundaries and data density, visualized the same way as in Figure 1. 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 (reference vectors) are divided among the classes in proportion to the sizes of the classes, similarly to the Conscience algorithm results in Figure 1: 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 intercluster gaps. Original figure is color. Paper posted at <http://www.ece.rice.edu/erzsebet/papers/esann04-2.pdf>.

respective pixel. See, for example, [4] 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.

We used 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 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. Figure 3 demonstrates the discovery of one such - and as it turned out very small - cluster. It also shows another small cluster (pale aqua, class V) that was known at the time of an earlier supervised classification, but was more definitely outlined by BDH clustering. Figure 4 compares the two SOMs. Shown on the left is the 40 x 40 SOM formed by BDH learning with $\alpha_{intended} = -0.8$, using only the upper right quadrant of the image (framed in Figure 3), *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. Also indicated are two other small clusters, that correspond to the previously known V (pale aqua) and B (white) classes from the supervised class map. 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 Conscious SOM in contrast to 7 PEs in the BDH SOM where it is also contoured by better developed “fences”. Similarly, the previously known small V class is represented by 4 PEs in the Conscious 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 right quadrangle, and as seen from the image on the right.

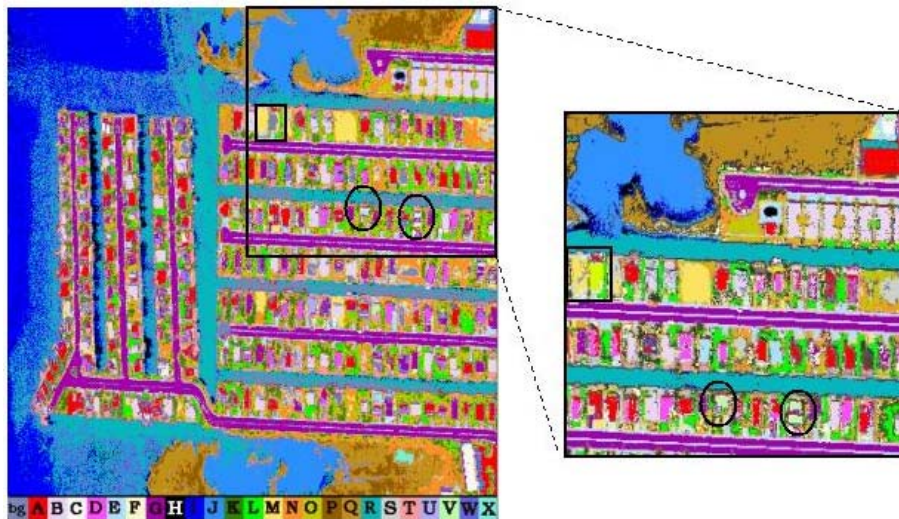


Figure 3: Comparison of supervised classification and BDH clustering with $\alpha < 0$. **Left:** An earlier supervised classification that satisfactorily mapped 24 known cover types of interest. Shown centered in the small black rectangle within the framed upper right quadrant is an unclassified grey spot (the color of the background, 'bg') apparently of the shape of a building, to the right of a yellow rectangular patch. **Right:** SOM clustering using BDH magnification control with $\alpha = -0.8$ on the upper right quadrant of the image. First, notice that the agreement between the supervised class map and this cluster map is striking, which inspires confidence in the clustering. Secondly, notice that the spot that remained unclassified in the supervised map is now filled exactly and with a color (greenish-yellow) that is different from all previous class colors and its spectral signature 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 4 shows the SOM view of this discovery. Original figure is in color. Download paper from <http://www.ece.rice.edu/erzsebet/papers/esann04-2.pdf>.

The previously known 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 SOM the BDH performed negative magnification.

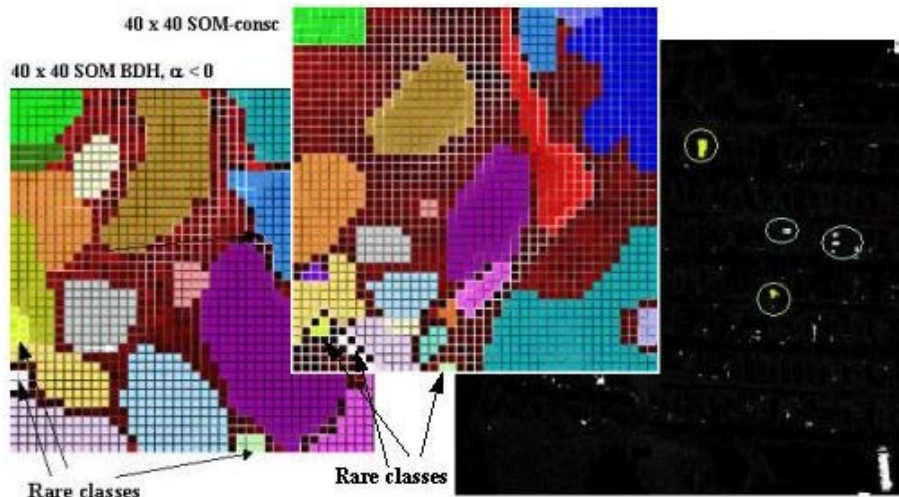


Figure 4: Comparison of SOMs developed by BDH versus Conscience learning. **Left:** The SOM learned by BDH, $\alpha < 0$, using the upper right quadrant of the 512 x 512 pixel Ocean City image. **Middle:** The SOM learned by the Conscience algorithm ($\approx \alpha = 1$), using the entire Ocean City image. **Right:** The rare classes in the image. It is apparent, as explained in the text, that the rare clusters are magnified in the BDH SOM in comparison to the Conscience SOM. Original figure is in color. Download paper from <http://www.ece.rice.edu/erzsebet/papers/esann04-2.pdf>.

4 Conclusion and future work

We presented systematic experiments with the map magnification control by Bauer, Der, and Herrmann (BDH) [2], on data for which the BDH scheme is not supported by existing theory. Based on our observations of the systematic BDH behavior on 2-D “forbidden” data, we were able to induce $\alpha = 1$ mapping on 6-D data. We also showed that negative magnification worked on 8-D real image data and helped discover small clusters. While the range of our studies is too limited to draw conclusions, the simulations indicate consistent behavior of the BDH on “forbidden” data. This encourages further simulations to investigate the predictability of the BDH for potential analyses of complex, high-D data.

References

- [1] A. Jain, E. Merényi. Forbidden magnification? I. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'04)*, Brussels, Belgium, 2004. D facto publications.
- [2] H.-U. Bauer, R. Der, M. Herrmann. Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, 8(4):757–771, 1996.
- [3] D. DeSieno. Adding a conscience to competitive learning. In *IEEE Int'l Conference on Neural Networks*, volume 1, pages 1–117–124, 1988.
- [4] T. Villmann, E. Merényi, and B. Hammer. Neural maps in remote sensing image analysis. *Neural Networks*, 16:389–403, 2003.