An application of the Self-Organizing Map and interactive 3-D visualization to geospatial data.

Masahiro Takatsuka
GeoVISTA Center, Dept. of Geography, The Pennsylvania State University, University Park, PA 16802 USA
masa@psu.edu, http://www.geovistastudio.psu.edu/

Abstract. Computer technologies have been rapidly improving throughout the last couple of decades, and they are now at the stage of allowing scientists to carry out data analyses that deal with very complex and multivariate datasets. Moreover, there are growing numbers of researchers who wish to carry out such tasks in real-time. Traditional data analyses and visualization techniques are useful but not sufficient to achieve those tasks. The Self-Organizing Map (or Kohonen's Feature Map) is one of the many modern data analysis tools that researchers have found useful in analyzing high-dimensional (multivariate) datasets such as atmospherical and demographical data. It is often used for such data analyses because of is multidimensional scaling and topological mapping capabilities. However, information loss caused by multidimensional scaling sometimes results in difficulty in interpreting an SOM when it is visualized in 2-D space. This study presents the use of the SOM for geospatial data analysis with the help of Java-based advanced 3-D visualization tools and a visual programming environment (GeoVISTA *Studio*) in order to gain deeper understanding of those complex datasets.

1. INTRODUCTION

Most of us have witnessed the advances in modern computing over the last three decades (Ceruzzi, 1998), and it seems that it keeps improving its capabilities and performance with the help of semiconductor technology (Geppert, 1999). This dramatic improvement in computing has changed how geographic and other types of massive data are collected and/or processed.

Decennial census data collected by the U.S. Census Bureau contain a colossal amount of demographic information. There are hundreds of census variables available to analyze various demographic phenomena. Those who study a particular phenomenon normally select appropriate census variables according to a theory or hypothesis of the model. Even though an analyst may successfully choose variables suited to analyze the phenomenon, the number of parameters, which might be contributing to the phenomenon, is often still enormous. Hence data analyses using the power of modern computing technology would help to uncover various insights into the phenomenon.

Back in 1997, it was predicted that there would be at least 31 satellites in orbit (landsat-like, high resolution, hyperspectral and radar) capable of producing massive data of 30 meters or better resolution (Stoney, 1997). The amount and quality of the remote sensing information provided by these satellites will have significant impact on our knowledge and understanding of the Earth, providing better data analysis techniques which are capable of handling these massively complex datasets.

In order to handle such complex and high dimensional data, the demand for more sophisticated data analysis techniques, including information visualization, has been increased. The modern advances of computing and visualization technologies (Brown et al., 1995) also allow us to implement and to utilize such computing-intensive, sophisticated data analysis techniques (Schwan, K. et al., 1995; Hecht-Nielsen, 1990). Moreover, there are now great demands for carrying out such tasks in real time in order to test many hypotheses with different parameters allowing analysts to quickly gain insights into complex models (Schwan, K. et al., 1995).

The Self-Organizing Map is one emerging Artificial Neural Network technology that has been found useful to analyze massively complex datasets. The applications of the SOM can be found not only in the fields of engineering but also in other areas such as the medical, agricultural and social science fields (Kohonen, 1995; Tokutaka et al., 1999). This paper discusses an application of the SOM to geographic (demographic) datasets with the help of interactive 3-D visualization.

2. DATA ANALYSIS USING NEURAL NETWORKS

Neural networks are one of many emerging computing technologies that have been actively studied over the last three decades (Hecht-Nielsen, 1990). They are inspired by ideas from neuroscience that a sophisticated computing system can be constructed from a network of simple processing units (e.g., neurons). How neural networks work depends on the interconnectivity between neurons. In some neural networks, the connections can be pre-computed from equations of models. However, most neural networks exploit their learning capabilities to obtain appropriate connections from sample data.

An artificial neuron itself carries out very simple signal processing using its internal function, which is usually a non-linear function such as a sigmoid function. Due to this non-linear nature of neurons, a massively connected network of neurons can capture very complex and highly non-linear characteristics of data (Bigus, 1996; Fayyad, 1996).

When neural networks are used to capture complex structural information of the feature space, it is often necessary to analyze what the networks have learned or discovered in addition to just using them to obtain answers for unknown input data. However, this is one of the most challenging tasks of the neural networks (Bigus, 1996).

There are different ways to carry out this task. One approach is to use the neural network as a "black box" and to make educated guesses by monitoring its' output responses against the controlled input data. This type of analysis is called *sensitivity analysis* (Bigus, 1996). Assume that a neural network is trained with a multivariate dataset of census data, which represent the cultural and economic variables, in order to analyze the phenomenon of gentrification. One can investigate what the neural network has learned by examining the impact that a particular input variable has on the gentrification. This is normally done by monitoring the changes in outputs while varying only the particular input variable and fixing the other input variables. The difficulty of this type of analysis is that one has to exhaustively repeat this process for each variable until relationships between the input variable and the gentrification are obtained.

Another approach is to transform the state of the neural network (what the neural network has learned) into more human readable forms. For example, *if-then* rules, which are typical knowledge representations in many standard artificial intelligence techniques, are very intuitive and easy to analyze by humans in order to find the importance of some particular input variables on the output. This *rule generation* has been studied during the last couple of decades (Gallant, 1988; Kane and Milgram, 1994). However, most of the rule-based knowledge representations rely on a binary conditioning (Russell and Norvig, 1995). Hence it is sometimes very difficult to map a neuron's nonlinear state onto the binary state.

A third approach, which this study has taken, is to graphically illustrate the internal state of the neural network allowing the human to visually inspect what the network has learned. There are many ways to visualize the neural networks. The simplest method is to use standard 2-D or 3-D numerical visualization techniques such as histograms, scatter plots and surface plots. However, they often only describe the overall behavior of the network or overall qualitative characteristics of the feature space. It is very difficult to analyze locally how each neuron describes a part of a feature space. Another method is to directly illustrate the state of neurons in the form of a network. A network graphic and the *Hinton diagram* belong to this category. Although this type of method allows you to see the state of an individual neuron or a connection weight, it is still not straightforward enough to clearly associate them with the original high dimensional feature space.

In this study, a Self-Organizing Map (SOM) is used to analyze a demographic dataset. As will be shown, the SOM captures the structural information of the feature space using the multidimensional scaling principal while preserving topological information as well. This is a great advantage in visualizing the state of neurons and connection weights over other types of neural networks.

3. BASICS OF A SELF-ORGANIZING MAP

The SOM is a set of artificial neurons, which are ordered in $\boldsymbol{N}^n$ space. A two dimensional array (n=2) is the most common map and is used to map an input signal in $\boldsymbol{R}^m$ (m>n) space onto the two-dimensional space (see Figure 1:).

An SOM typically consists of two layers. One is an input layer into which input feature vectors will be fed and the other layer is a two-dimensional competitive layer, which orders the neurons' responses spatially. Neurons can be arranged on a rectangular map so that they can be implemented using a simple 2-D data array. A hexagonally arranged neuron map is, however, often used because it has the advantage that Euclidean distances between adjacent neurons are equal for all six nearest neighbor neurons (Kohonen, 1995).

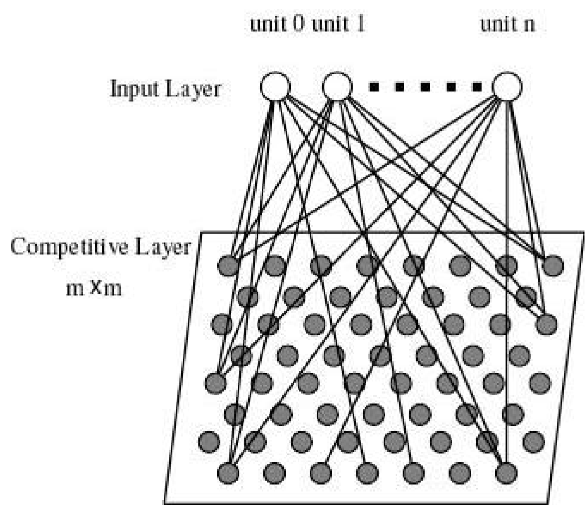

Figure 1: A hexagonal array of neurons in the SOM.

3.1 Learning Algorithm (Kohonen, 1990)

The SOM is trained without teacher signals (unsupervised), unlike some other ANNs in which supervised training is used, such as in backpropagation networks. The learning algorithm used in this study is the same as Kohonen's algorithm (Kohonen, 1989) and described as follows:

Let $\boldsymbol{x}_i = \{x_1, x_2, \ldots, x_n\}^T \in \boldsymbol{R}^n$ be the i-th input feature vector, which will be fed into the neurons in the input layer. Because the task of the input neurons is just to pass the input feature vector onto the competitive layer, the input vector $\boldsymbol{x}_i$ will be used as an output vector of the input layer. The neurons in the input layer and the neurons in the competitive layer are fully connected and the connection weights of the competitive neuron j shall be denoted by $\boldsymbol{w}_j=\{w_{j1}, w_{j2}, \ldots, w_{jn}\}^T$. $\boldsymbol{R}^n$. The j-th competitive neuron computes the similarity of $\mathbf{x}_i$ and $\boldsymbol{w}_j$ as the output y_j:

$$y_i = \left\| x_i - w_j \right\| = \sqrt{\sum_{k=1}^{n} \left(x_i - w_{jk} \right)^2} \ .$$

There are several ways to measure the similarity (Kohonen, 1989); the Euclidean distance is used in this study. By training the connection weights according to the following algorithm, the competitive neurons will have the appropriate weights so that the competitive neurons adjacent to each other will respond to input vectors that are close in the input feature vector space.

First, all competitive neurons compute the match of the input $\boldsymbol{x}_i$ with their connection weights $\boldsymbol{w}_j$. Then, the best-matching neuron and its adjacent neurons now update their connection weights with the following Hebbian learning rule:

$$\boldsymbol{w}_j(t+1) = \boldsymbol{w}_j(t) + .(t).(t)[\boldsymbol{x}_i(t) - \boldsymbol{w}_j(t)],$$

where t is a variable in the discrete time index, $.(t)$ is a learning rate and $.(t)$ is a neighborhood kernel. The learning rate is a monotonically decreasing function of time ($0 < .(t) < 1$) and defined as:

$$.(t+1) = .(t)(1 - t/T),$$

where T is a training period. The neighborhood kernel defines the adjacent region of the best match neuron c. All neurons within this region will update their connection weights. The neighborhood kernel is also a monotonically decreasing function of time:

$$h(t) = \begin{cases} 0 \longrightarrow \left\| r_j - r_c \right\| \leq N_c(t) \\ 1 \longrightarrow otherwize \end{cases} ,$$

$$N_c(t+1) = N_c(t)\,(1 - t/T),$$

where $\boldsymbol{r}_j$ and $\boldsymbol{r}_c$ are the position vectors of the j - th and the c - th neuron, respectively, in the two-dimension array so that.$\boldsymbol{r}_j$ - $\boldsymbol{r}_c$.represents the Euclidean distance between those neurons in the array. $N_c(t)$ is the radius of the kernel which actually defines the kernel region. After this learning process, the neurons that are ordered in the two-dimension array preserve the topological information of the original feature space. Neurons geometrically close to each other in the map will represent the input features, which are close to each other in the input feature space.

In practice, the training data $\boldsymbol{x}_i(i = 1, \ldots, n)$ are iteratively used during the learning period T. After the training, input feature vectors, which have continuous values as elements are topologically mapped onto the two-dimensional map. This SOM is used in order to analyze segmented surface parts of objects. In addition, it transforms the feature vectors of surface parts into two-dimensional vectors that represent the positions of neurons.

4. VISUALIZING AN SOM

The SOM is typically trained to map a high dimensional real space to a 2-D integer space. Once the SOM is trained, all samples used in training will be represented by a neuron with equal probability. This is one of the SOM's mathematical goals and one can use the trained SOM to model the probability density function of the given sample. Other uses of the SOM are multidimensional scaling (data encoding), cluster discovery, and visualizing high dimensional data space.

4.1 2-D Visualization

When the SOM is to discover some structure of the given samples in the feature space, it often is useful to visualize the finding in the form of cluster formation.

The typical 2-D SOM is used to achieve multidimensional scaling and to map the higher dimensional space onto a 2-D space. Once the high dimensional data is mapped onto the 2-D space, it is very intuitive and easy to analyze the data structure of the samples. Moreover, the SOM offers the topological information of the samples in the feature space in the form of adjacency network of neurons in the 2-D space.

By exploiting these features of the 2-D SOM, visualization techniques to depict the data structure of the feature space in the form of clustering of neurons in the SOM have been developed (see Ultsch, 1993; Kraaijveld et al., 1992). This visualization typically uses a gray scale to illustrate the distance between connection weights. The light shading typically represents a small distance and the dark shading represents a large distance (see Figure 2: for an example).

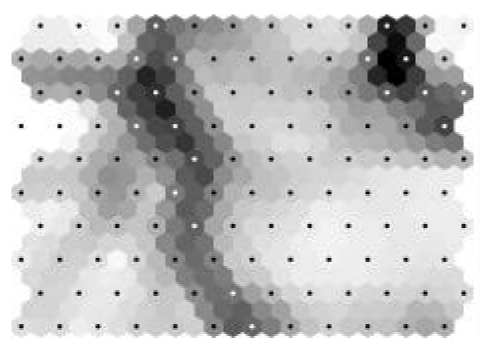

Figure 2: An example of a gray scale distance map representing a 12x10 SOM. A cluster can be identified as a region surrounded by dark regions.

This type of visualization is useful as long as relatively clear cluster boundaries exist or the granularity of the distance differences is large. When the cluster boundaries get fuzzy or the granularity of the distances becomes too small to represent with a gray scale, it starts getting difficult to see fuzzy cluster landscapes in the gray scale. Moreover, since all distance values are normalized, only relative (qualitative) analysis is allowed. Subsequently, this "gray scale distance map" cannot be used to compare different SOMs' mapping results.

4.2 3-D Visualization

In this study, the state of the SOM is visualized in a 3-D space. The distances between connection weights are used to compute height values (along the z axis) and color values. The color serves the same visualization effect as that of the gray scale distance map. However, by using the height, the distances no longer need to be normalized. Hence they represent the real Euclidean distances in the original feature space, and can be used to compare different cluster formation among different SOMs. Moreover, it is much easier to see clusters in the 3-D surface form and is more intuitive to interpret the distances between clusters. One can easily perceive the distances in the original feature space as geodesic distances on this synthetic 3-D surface.

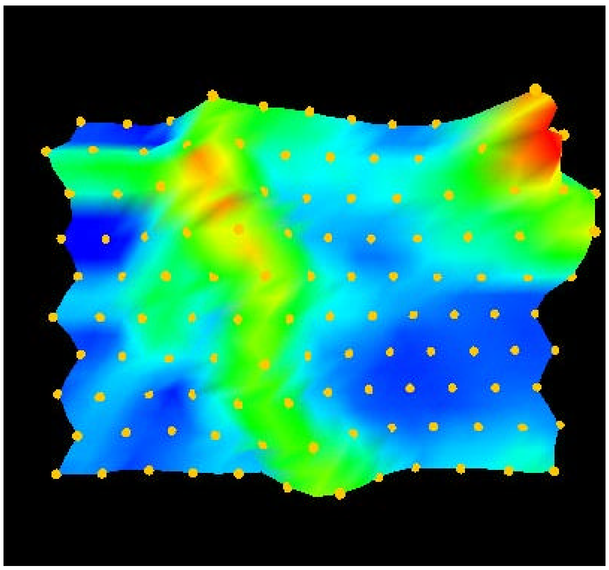

Figure 3: An example of a 3-D distance map representing the same SOM as Figure 2:

Figure 3: is an example of a 3-D version of the distance map. It represents the same 12x10 SOM as Figure 2:. Neurons are indicated by the orange dots. In this figure, the color is used to represent normalized distances...Java3D™ linearly interpolates colors between vertices along surfaces. Since colors were assigned to the *contiguous 3-D surfaces* unlike the gray scale distance map, the color gradation reveals underlying 3-D surface structures. Consequently, one can observe cluster formation more clearly. For instance, two clusters can easily be identified at the bottom-left corner in Figure 3:, whereas it is very difficult to distinguish these two regions in Figure 2:, even though there is a slight difference in gray scale colors. Furthermore, one can identify two clusters in the top-middle region in the 3-D distance map but it is difficult to see them in the gray scale distance map. These observations can be confirmed by looking at the 3-D surface from a different viewing angle. The clusters are clearly indicated by the ridges in Figure 4:.

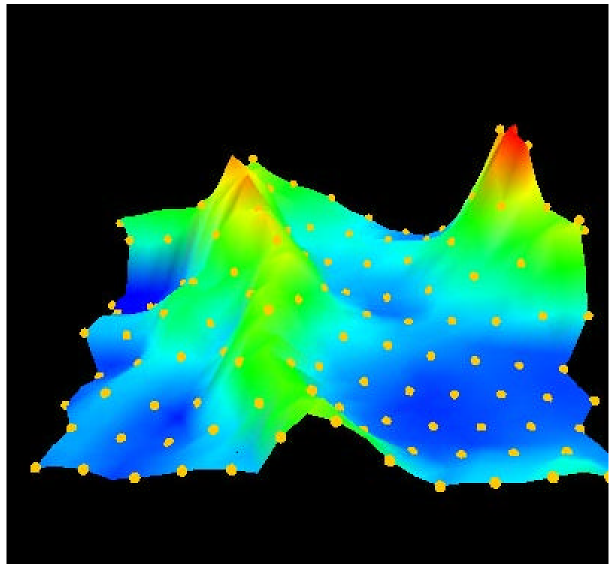

Figure 4: The 3-D distance map from a different viewing angle. Notice the clusters surrounded by ridges.

4.3 Dynamic visualization

Here, the word *dynamic* does not mean a simple animation. It means that the contents of the visualization can be manipulated and modified in real time (for instance, in the form of steering). The integration of steering and visualization has been recognized as a very important aspect of studying complex models and datasets (Schwan, K. et al., 1995).

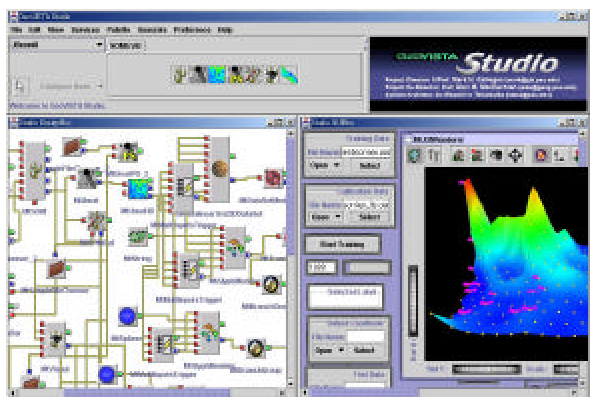

Figure 5: A screenshot of GeoVISTA *Studio*

In this study, the visualization/data analysis program is designed and constructed using GeoVISTA *Studio* (Takatsuka and Gahegan). *Studio* is a coding-less visual programming environment based on Java™ and JavaBeans™ architecture. A user builds a program by connecting various computing modules (see Figure 5:). In *Studio* the design phase and the execution phase are merged and the program being designed is always alive. Hence a user can dynamically configure interconnection between modules. Consequently, the user can explore numerous visualization possibilities by assigning different data values to different visual variables. Moreover, the 3-D visualization modules in *Studio* are based on Java3D™ technology, which allows a user to update the 3-D scene in real time by supplying numerical values, which are computed by the model steering modules.

In the late 80's and the early 90's many scientific visualization tools exemplified there were two types of scientific visualization tools. One was to provide simple but fast tools that enable a user to quickly display different views of the data. The other was to provide tools that are more computationally intensive but capable of generating much richer scenes (Saltzman, J., 1990). In either case, a user has to combine various different software and/or hardware tools in order to manipulate data and to obtain visualization results. Tools like GeoVISTA *Studio* allow a user to combine various computing modules easily and effectively in order to accomplish sophisticated scientific visualization (Takatsuka and Gahegan).

In this environment, the training process of the SOM can be visualized in real-time. One can observe how neurons are trained and spread across the feature space by monitoring the 3-D visualization of the current status of the SOM. It might require some training to make sense of this dynamic visualization, but it is useful for analyzing how particular ordering of the training samples have influence on the training process.

5. EXAMPLE

In this section, an example of demographic profile analysis is described. It demonstrates the use of the SOM and the interactive 3-D visualization to gain an understanding of the gentrification phenomenon at Harrisburg, PA.

5.1 Datasets

The various census data were collected for the twenty-three tracts of the City of Harrisburg (See Figure 6:) as listed in Table 1 (Mostert, 2001). These data were chosen to reflect the city attributes (such as class shift, housing stock, suburban expansion, economic conditions, lending institutions, government involvement, etc) in the development of gentrification (Lakshman, 1996). The raw census values were converted into percentage of population variables, and monetary values (Income, Rent and Value) were adjusted for inflation to 1999 dollars based on the consumer price index (CPI) (U.S. Department of Labor Bureau of Labor Statistics 2000). Even though these census data describe microscopic changes, they are sufficient for an indication of neighborhood level temporal shifts in class and value.

Table 1: Census Variables (Mostert, 2001).

Census Variables
Total Population
Young (20-34)
Middle Aged (35-54)
Married Couples with Children under 18 years
Single
Married
White Race
Black Race
All Housing Units
Owner-Occupied Housing Units
Renter-Occupied Housing Units
Vacant Housing Units
Persons 25 years and over
Completed High School
College 1-3 Years
College Graduates
Worker's 16 years and over
Worked in SMSA of residence
Harrisburg City
Remainder of Dauphin County
Worked outside SMSA of residence
Employed persons 16 years and over
Civilian Labor Force Unemployed
Executive, administrative, and managerial occupations
Administrative support occupations, including clerical
Operatives (including machine operators, assemblers, and inspectors)
Females 16 years and over
Females in the Civilian Labor Force Employed
Raw Monetary Values
Median Family Income (dollars)
Median Gross Rent (dollars) or Median (dollars) Rent Renter-Occupied Housing Units
Median (dollars) Value Owner-Occupied Housing Units

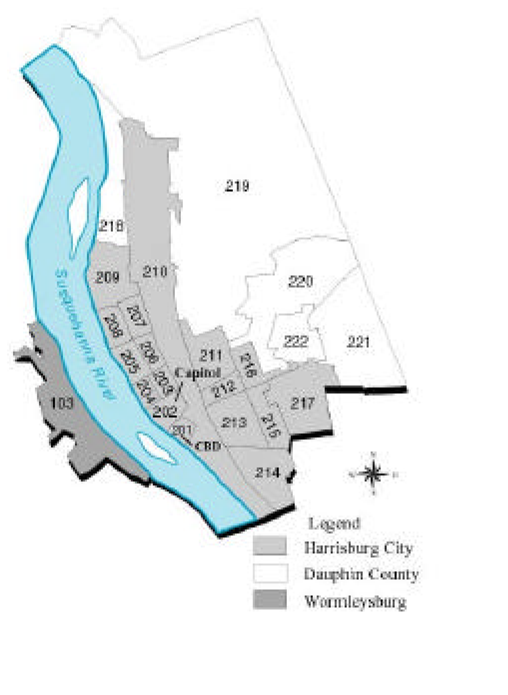

Figure 6: Harrisburg, Wormleysburg, and selected Dauphin County, Pennsylvania Census Tract Boundaries.

5.2 Program Design

The program for analyzing these gentrification data was developed using GeoVISTA *Studio*. Figure 7: shows the design of the analysis program. It contains three major parts. There is an SOM section at the top-left region. There are three main components in this section. One is the SOM component that trains the SOM. The SOM calibration component labels each neuron according to the calibration data that contain labels for known samples. The last component finds a corresponding neuron for a given sample data and computes the classification errors.

The second section (in the top-right region) contains components for 3-D visualization. In this section, the information of the trained SOM is transformed into a synthetic 3-D surface using x and y coordinates of neurons and distance values of neurons' connection weights. The distance values were used for height values of the synthetic surface. Hence the high elevation surface indicates the great separation between neurons in the feature space.

The last section contains modules to create Sammon mapping (Sammon, JR., 1969) from the trained SOM. This Sammon mapping module computes the location of neurons in the 3-D space from their connection weights. Sammon mapping has been used to visualize *N*-dimensional data in two or three-dimensional space in order to discover some structural information of multidimensional data. This mapping illustrates a more realistic view of data structure in the original feature space. Although it was not included in the experimental results, it is useful to use this mapping for assisting the interpretation of the 3-D distance map.

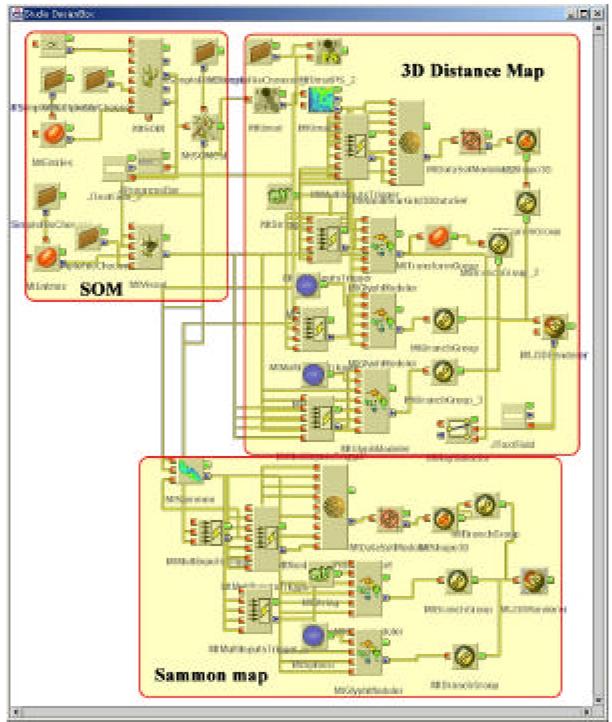

Figure 7: A Studio design for 3-D Visualization of the SOM.

5.3 The SOM training parameters

The size of the SOM used in the experiment was 30x30. An SOM of size 20x20 would normally produce the similar cluster formation, but the size of 30x30 produces more detailed 3-D surfaces. The training of the SOM was carried out in two phases. The training parameters are listed in Table 2. During the first stage of training, the SOM is trained with the large initial learning rate, the neighborhood size and the short learning step. The purpose of this stage is to roughly place neurons in the feature space. During the second stage, the locations of neurons are fine tuned by using the much smaller initial learning rate and neighborhood size.

Table 2: Training parameters for the SOM.

	Initial Learning Rate a	**Neighborhood size N_c**	**Learning Steps**
1st Stage	0.5	0.8 * map size	1000
2nd Stage	0.2	0.2 * map size	10000

5.4 Chronological cluster analysis

In order to find the movement of gentrification in each decade, three SOMs were trained with 1970, 1980 and 1990 data. Figure 8: shows the 3-D distance map for 1970. It is clear that the capitol (202) and the center business district (CBD) (201) are located at the top-left corner with a relatively large distance from other inner city tracts (203 and 208) and tracts at the edge of suburban (215 and 217). By judging the height of cluster boundaries, the clusters are organized on the map from the inner city tracts to the suburban tracts in a counter clockwise direction.

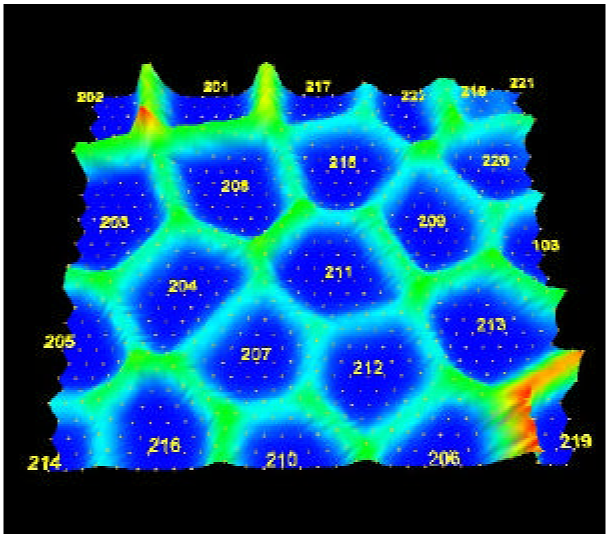

Figure 8: The 3-D distance map for 1970

There were several significant events, which took place between 1970 and 1980. There was the Hurricane Agnes flood, the urban renewal projects, and the establishment of the Harrisburg Historical Association and the Shipoke Historic District (U.S.G.S, 1996; Redevelopment Authority, 1967, 1971, 1973a, 1973b; City of Harrisburg, 1998). The tracts (204, 205, 208 and 210), which are known to be affected by these events, showed significant movement in the SOM. In the 1980's 3-D distance map (Figure 9:), the majority of tracts seem to be organized in a clockwise direction from top-left (inner city) to bottom-left (suburban). However those four tracts were placed in the middle of the map away from their neighboring tracts.

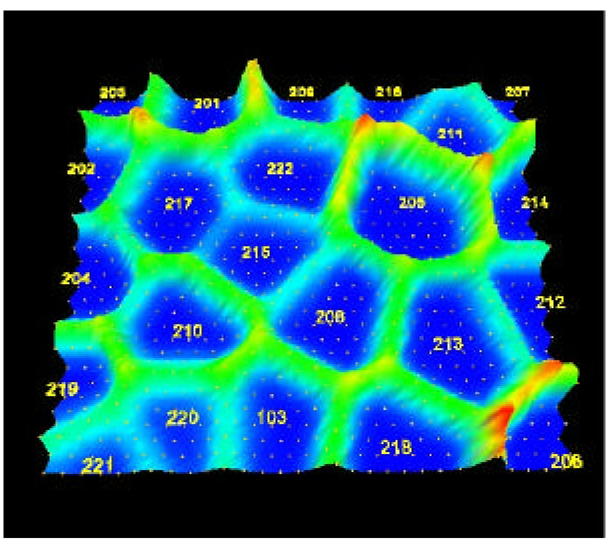

Figure 9: The 3-D distance map for 1980

In the 1990's map (Figure 10:), tracts for the suburban area are located at the top-left area, inner city tracts are at the top-right corner (except capitol 201), middle class tracts (209 and 215) are at the top-middle and blue collar are at the bottom half of the SOM. Since those inner city tracts and middle class tracts are situated close together near the suburban tracts, this could be an indication of the gentrification.

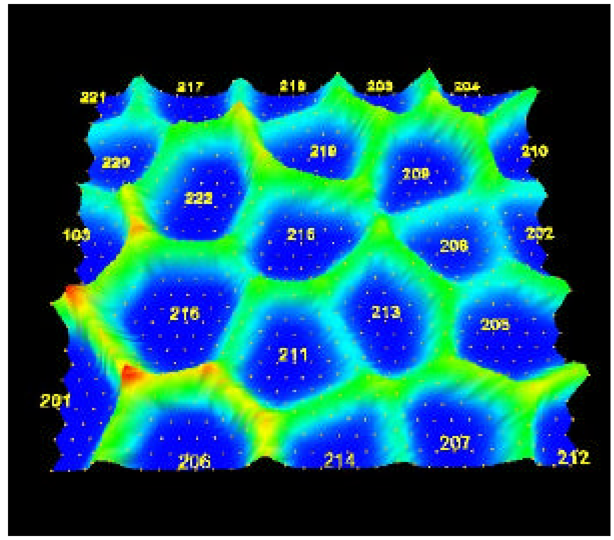

Figure 10: The 3-D distance map for 1990

5.5 Temporal cluster analysis

In this section, the movement of each tract over three decades is analyzed in a single SOM. For this experiment, all data from 1970, 1980 and 1990 were used to train the SOM to find data structure (clusters) of census data. Once the SOM was trained, it was labeled using separate datasets from the '70's, '80's and '90's as shown in. Figure 11: ~ Figure 13:.

All the tracts were located in order from bottom-left (inner city tracts) to top-left (suburban tracts) for the '70's data (Figure 11:). This is consistent with the result shown in Figure 8:, indicating the status of the inner city is very different from those of suburban and middle class tracts.

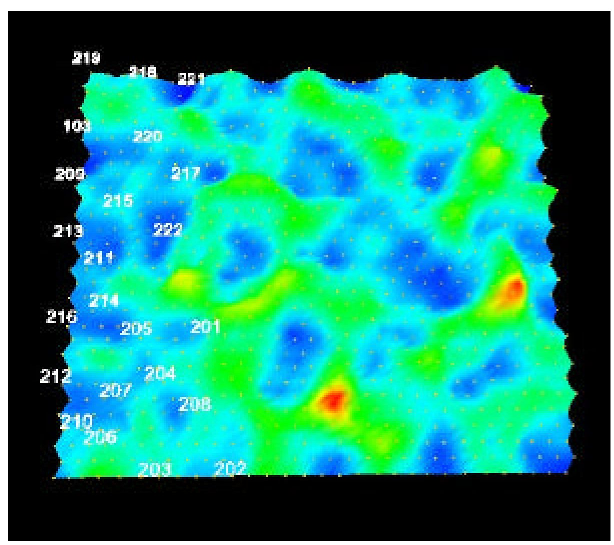

Figure 11: The 3-D distance map for three decades with 1970 tact labels.

In 1980, those suburban tracts moved towards the middle of the map and those supposed to be affected by various events in the 1970's are located at the middle-bottom of the map. Those belonging to middle class and blue collar moved to the right-middle region except tract 206, which did not show significant change. However, as is shown in Figure 13:, tract 206 seems to follow the movement of tract 207. Some inner city tracts, such as 202, 204, 205 and 208 are now closer to suburban and middle class tracts.

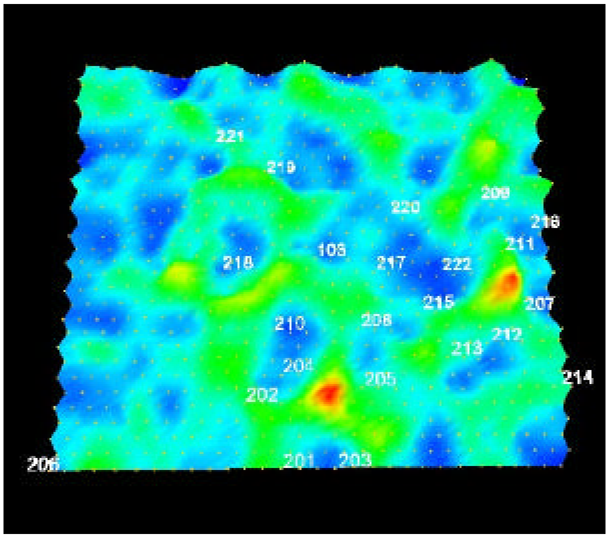

Figure 12: The 3-D distance map for three decades with 1980 tact labels.

In the figure for 1990, the inner city tracts were located at the top-right corner with the middle class and suburban tracts in their vicinity. They are now much closer to middle class and suburban tracts than blue-collar tracts. However, there are still differences between them as high ridges between them indicate. Since it is known that the tracts 203, 204 and 210 have been gentrified, it is possible to assume that nearby tracts (202, 205 and 208) are currently experiencing the gentrification. It is also interesting to see how the tract 206 will develop (either move towards the blue collar tracts or become gentrified).

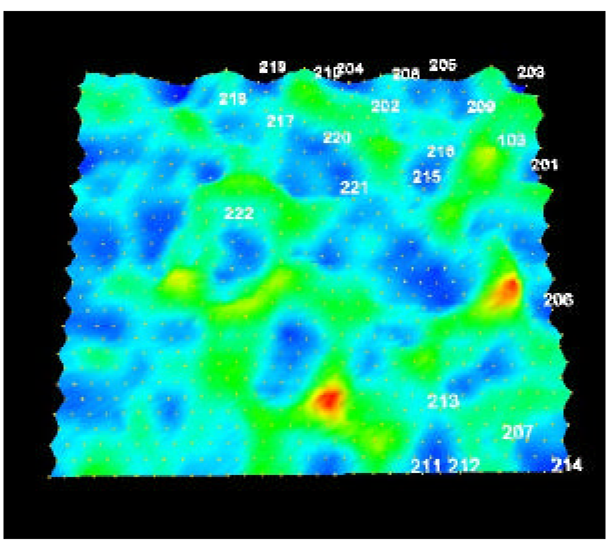

Figure 13: The 3-D distance map for three decades with 1990 tact labels.

It should be noted that all visual aspects of the 3-D distance map could be customized in real time. For instance, one might want to increase the height resolution of the map in order to emphasize the structural details of the 3-D surfaces. Another example of interactivity would be to assign different numerical values to visual variables. For example, it is possible to create a 3-D surface using quantization errors of neurons instead of distance between them. Moreover, the SOM and 3-D visualization components can be connected to other computational (e.g. statistical) or visualization (traditional map and PCP) components in order to provide a more different view of the data.

6. CONCLUSION

This paper showed how the SOM could be applied to analyze a complex geographic (demographic) dataset. In addition, it demonstrated the use of 3-D visualization to depict the state of the SOM more clearly than the conventional gray scale distance map.

Recent geographic data have become more complex and too large to be handled by the traditional statistical tools. This study utilized the multidimensional scaling and topological mapping capability of the SOM to analyze high dimensional (26 dimension) demographic data. In order to learn what the SOM have extracted from this high dimensional data space, the 3-D version of the distance map was used. The use of 3-D visualization allowed a clear illustration of the underlying data structures, which were more difficult to see in the conventional gray scale distance map.

This study demonstrated the use of these tools to analyze the gentrification in Harrisburg, Pennsylvania. By analyzing the clusters found by the SOM or by labeling neurons to track the movements of census tracts over a period of time, the gentrification phenomenon has been observed.

In further research, these tools must be linked to different types of data analysis tools in order to assist the extraction of more information from the data.

ACKNOWLEDGEMENT

I would like to thank Christian Mostert for his work in preparing the census data and for his help in analyzing the gentrification in Harrisburg, PA.

REFERENCES

Bigus, J. P. (1996) *Data Mining with Neural Networks*, McGraw-Hill, NY

Brown, J. R. et al. (1995) Visualization: Using Computer Graphics to Explore Data and Present Information, John Wiley & Sons, Inc. NY.

Ceruzzi, P. E. (1998) A History of Modern Computing, MIT Press, Cambridge, Mass.

City of Harrisburg (1998) "Revised Draft: Zoning Code, Title Seven - Part 300 of the Codified Ordinance of the City of Harrisburg." Trans.

Fayyad, U. M. (ed.) (1996) *Advances in knowledge discovery and data mining*, MIT Press: AAAI Press: Menlo Park, CA

Gallant, S. I. (1998) "Connectionist expert systems", *Communications of the ACM*, vol. 31, no. 2, pp. 152-169

Geppert, L. (1999) "The 100-million transistor IC", IEEE Spectrum, vol. 36, No. 7, pp.22-24.

Hecht-Nielsen, R. (1990) *Neurocomputing*, Addison-Wesley, Reading, Mass.

Kane, R. and Milgram, N. (1994) "Financial forecasting and rules extraction from trained networks", *Proceedings of IEEE International Conference on Neural Networks*, vol. 5, pp. 3190-3195

Kohonen, T. (1989) *Self-Organization and Associative Memory*, Springer-Verlag, NY

Kohonen, T. (1990) "The Self-Organizing Map", *The Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464-1480, September.

Kohonen, T. (1995) *Self-Organizing Maps*, Springer-Verlag: Berlin

Kraaijveld, M. A. et al. (1992) "A non-linear projection method based on Kohonen's topology preserving maps", *Proceedings of the 11th International Conference on Pattern Recognition*, pp. 41-45

Lakshman, Y. (1996) "What Causes Poverty? A Postmodern View", Annals of the Association of American Geographers, vol. 86, No. 4, pp. 707-728.

Monstert, C. A. (2001) Applications of Exploratory Analysis Techniques: A Study of Gentrification in Harrisburg, Pennsylvania, Master Thesis, College of Earth and Mineral Science, The Pennsylvania State University.

Redevelopment Authority of the City of Harrisburg (1967) "Survey and Planning Application for the Hamilton Urban Renewal Area in Harrisburg, PA." Trans. Harrisburg: Produced by the David Walker Associates, Inc. Urban Renewal Planning and Development Consultants.

Redevelopment Authority of the City of Harrisburg (1971) "Survey and Planning Application for the Hamilton Urban Renewal Area in Harrisburg, PA." Trans. Harrisburg: Produced by the David Walker Associates, Inc. Urban Renewal Planning and Development Consultants.

Redevelopment Authority of the City of Harrisburg (1973a). "Cameron-South Urban Renewal Project, Loan and Grant Application." Trans. Harrisburg.

Redevelopment Authority of the City of Harrisburg (1973b) "Penn-Susquehanna Urban Renewal Project, Disaster Project R-634(c) Draft parts I and II, Loan and Grant Application." Trans. Harrisburg.

Russell, S. J. and Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, NJ.

Saltzman, J. (1990) "Visualization Techniques for High Dimensional Data Sets", *Performance Instrumentation and Visualization*, Eds. Simmons, M. and Koskela, R., ACM Press: NY.

Sammon, JR., J. W. (1969) "A Nonlinear Mapping for Data Structure Analysis", *IEEE Transactions on Computers*, vol. C-18, No. 5, May.

Schwan, K. et al. (1995) "Integrating Program Steering, Visualization, and Analysis in Parallel Spectral Models of Atmospheric Transport", *Science Information Systems Newsletter*, issue 36.

Stoney, W. E. (1997) "REMOTE SENSING IN THE 21ST CNETURY: OUTLOOK FOR THE FUTRE", The Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Singapore, August 7.

Takatsuka, M. and Gahegan, M., "GeoVISTA *Studio*: A Codeless Visual Programming Environment For Geoscientific Data Analysis and Visualization", The Journal of Computers and Geosciences (in print).

Tokutaka, H., et al. (1999) "Application of Self-Organizing Maps (SOM) to Chemical Data Analysis", *Journal of Surface Analysis*, vol. 5, no. 1, pp. 102 - 105

Ultsch, A. (1993) "Self Organized Feature Maps for Monitoring and Knowledge Acquisition of a Chemical Process", *Proceedings of the International Conference on Artificial Neural Networks* (ICANN93), pp. 864-867

U.S. Department of Labor Bureau of Labor Statistics (2000) *Consumer Price Index for All Urban Consumers – (CPI-U) U.S. City Average All Items*, Available: ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt

U.S. Geologic Survey (1996) "Statewide Floods in Pennsylvania, January 1996: Fact Sheet 103-96." Trans.: U.S. Department of the Interior.

van Hulle, M. M. (2000) *Faithful Representations and Topographic Maps*, John Wiley & Sons, Inc.: NY