CAVE-SOM: Immersive Visual Data Mining Using 3D Self-Organizing Maps

Dumidu Wijayasekara, Ondrej Linda, Milos Manic

Abstract— **Data mining techniques are becoming indispensable as the amount and complexity of available data is rapidly growing. Visual data mining techniques attempt to include a human observer in the loop and leverage human perception for knowledge extraction. This is commonly allowed by performing a dimensionality reduction into a visually easy-to-perceive 2D space, which might result in significant loss of important spatial and topological information. To address this issue, this paper presents the design and implementation of a unique 3D visual data mining framework – CAVE-SOM. The CAVE-SOM system couples the Self-Organizing Map (SOM) algorithm with the immersive Cave Automated Virtual Environment (CAVE). The main advantages of the CAVE-SOM system are: i) utilizing a 3D SOM to perform dimensionality reduction of large multi-dimensional datasets, ii) immersive visualization of the trained 3D SOM, iii) ability to explore and interact with the multi-dimensional data in an intuitive and natural way. The CAVE-SOM system uses multiple visualization modes to guide the visual data mining process, for instance the data histograms, U-matrix, connections, separations, uniqueness and the input space view. The implemented CAVE-SOM framework was validated on several benchmark problems and then successfully applied to analysis of wind-power generation data. The knowledge extracted using the CAVE-SOM system can be used for further informed decision making and machine learning.**

I. Introduction

DATA MINING is becoming more and more important and useful as the amount and complexity of required data processing in various fields are rapidly increasing [1]. The available massive datasets have created a high demand for faster and more efficient data analysis methods. Data mining techniques constantly evolve benefiting from theories and techniques from many fields, including pattern recognition, high performance computing, and data visualization [2].

The effectiveness of the data mining process can be further increased by including humans in the data analysis process [3], [4]. Visual data mining is the process of exploration, interaction and reasoning with abstract data using natural human perception. The users of visual data mining tools are allowed to incorporate human intelligence in the data mining process. This is achieved via utilizing visual perception and thus combining flexibility, creativity and general knowledge of the human brain with the computational and storage capability of modern computers [5], [6]. The desired outcome is a visual discovery of important patterns and trends in the data. This knowledge and information can then be utilized in informed decision making [7].

Since humans are inherently accustomed to perceiving 2D or 3D spaces, visual data mining must first employ tools for dimensionality reduction of the multi-dimensional datasets. Such dimensionality reduction should preserve important features from the original input space. Intuitively, the higher the dimensionality of the desired output space, the more important features can be preserved through the transformation. Thus, visual data mining performed in 3D space should be superior to visual data mining in 2D spaces, as it allows for preserving richer information from the original data.

In order to successfully leverage the capabilities of the human observer, complex visualization methods and display environments have been proposed for visual data mining in the past [1]. As 3-dimensional (3D) visualization technology is advancing, more and more 3D visualization techniques are becoming available to users. Such advances have lead to the creation of Cave Automatic Virtual Environment (CAVE). CAVE is an immersive visualization technology that allows users to completely immerse themselves in the virtual environment. In this way, the virtual environment can be approached in an interactive and natural way. The CAVE and other virtual reality environments have been used in immersive visual data mining effectively in the past [3]-[11].

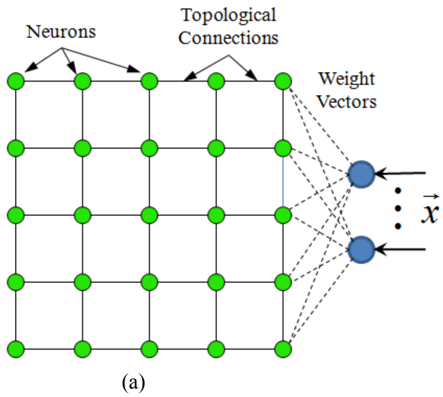
Artificial neural networks have been widely applied in the data mining field [12]. This can be attributed primarily to their capabilities to process, filter, model, and generalize based on multi-dimensional input datasets. The Self Organizing Map (SOM) algorithm is commonly used in visual data mining as dimensionality reduction and feature extraction tool [13]-[16]. SOM has the ability to translate features, trends, tendencies and spatial distributions existing in multi-dimensional data into spaces with lower dimensionality. In this manner, visualization of multi-dimensional data in human-perceivable spaces can be achieved. In addition, SOM can reduce the size of large datasets via mapping the input data onto a small set of neurons. Applications of SOM in visual data mining can be found in [17]-[20].

Various other methodologies such as PCA [21], Sammon mapping [22], Isomap [23], LLE [24] and manifold sculpting [25] are used in data mining for dimensionality reduction. However, the SOM was preferred in the presented work as a

Dumidu Wijayasekara is with the Computer Science Dept, University of Idaho, Idaho Falls, ID 83402 USA (phone: 208-533-8149; e-mail: wija2589@vandals.uidaho.edu).

Ondrej Linda is with the Computer Science Dept, University of Idaho, Idaho Falls, ID 83402 USA (e-mail: olinda@uidaho.edu).

Milos Manic is with the Computer Science Dept, University of Idaho, Idaho Falls, ID 83402 USA (e-mail: misko@ieee.org).



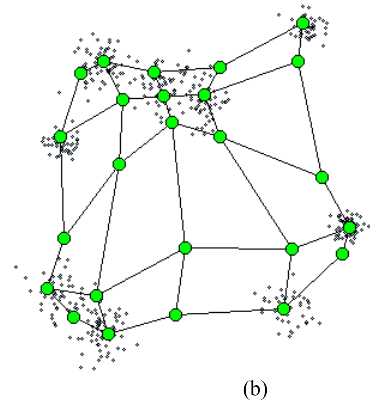


Fig. 1 Self-Organizing Map displayed in the output space (a) and in the input space adapted to 2D distribution of input points (b).

suitable algorithm for visual data mining due to the following reasons: i) fast convergence of the learning process, ii) ability to learn on-line, iii) ability to compress the input dataset into a set of neurons, and iv) ability to fix the number of output dimensions (e.g. 3 for the presented CAVE-SOM system).

This paper describes the development and implementation of a 3D SOM visualization developed for the CAVE system, with emphasis on interactive immersive visual data mining – the CAVE-SOM system. The primary advantage of using a 3D SOM over its traditional 2D implementation is the ability to preserve richer volume of information through the dimensionality reduction process. In order to allow the user to intuitively work with the complex 3D data visualization, the CAVE system is utilized. The CAVE-SOM provides a range of visualization modes, e.g. U-matrix [20], [26], histogram view [19], [27], connections [28], [29], separations, neuron uniqueness and input space view.

The developed CAVE-SOM system proved to be a powerful and intuitive visual data mining tool that can be used for visualization of high-dimensional large data sets. Furthermore, the CAVE-SOM system allows users to immersively interact and work with the data using advanced interfaces such as head tracking and a 6 Degrees-Of-Freedom (DOF) motion tracking tools. The proposed visual data mining framework was validated on benchmark datasets as well as on experimental data for wind power generation analysis.

The rest of the paper is organized as follows: Section II provides background review of the SOM algorithm. Section III introduces the CAVE visualization facility and the implemented CAVE-SOM data mining system. Section IV presents the experimental results, and Section V concludes the paper.

II. Self-Organizing Maps

The Self-Organizing Map (SOM) algorithm was developed in 1981 [30]. SOM uses unsupervised *winner-takes-all* competitive learning method together with cooperative adaptation to adjust itself to the topological properties of the input dataset. The SOM consists of a topological grid of neurons typically arranged in 1D or 2D lattice [31]. The fixed grid defines the spatial neighborhood of each neuron.

Each neuron maintains a synaptic weight vector $\vec{w} = \{w_1, \ldots, w_m\}$, where m is the dimensionality of the input space. The input dataset consists of input patterns that can be denoted as $\vec{x} = \{x_1, \ldots, x_m\}$. The structure of a 2D SOM is depicted in Fig. 1(a). All neurons are first randomly initialized and then iteratively adapted based on the training set of input data. The training process can be described in several steps as follows [31]:

Step 1 - Initialization: Randomly initialize all synaptic weight vectors in the input domain.

Step 2 - Sampling: Select a random input pattern $\vec{x}$ from the training dataset.

Step 3 – Competitive Learning: Find the Best Matching Unit (BMU) for the current input pattern $\vec{x}$. The BMU is found by minimizing the Euclidean distance between the input pattern $\vec{x}$ and the synaptic weight vectors $\vec{w}$:

$$BMU(\vec{x}) = \arg\min_j \left\| \vec{x} - \vec{w}_j \right\|, \quad j = 1, 2, \ldots, N \qquad (1)$$

Here, $BMU(\vec{x})$ is the best matching unit for input pattern $\vec{x}$, operator $\| \ \|$ denotes the Euclidian distance norm, and N is the number of all the neurons in the SOM.

Step 4 – Cooperative Updating: Update the synaptic weight vectors of all neurons in SOM using the cooperative update rule:

$$\vec{w}_j(n+1) = \vec{w}_j(n) + \eta(n)\, h_{j,BMU(\vec{x})}(n)\, (\vec{x} - \vec{w}_j(n)) \qquad (2)$$

Here, n denotes the iteration, $\eta(n)$ is the learning rate and $h_{j,BMU(\vec{x})}(n)$ is the value of the neighborhood function for the neuron j centered at $BMU(\vec{x})$.

Step 5 – Convergence Test: Until a specified convergence criterion is met go to **Step 2**.

The learning process is controlled by the dynamic learning rate and the neighborhood function. The neighborhood function is typically implemented as a Gaussian function centered at the selected wining neuron. Its amplitude applied to neuron j can be calculated as follows:

$$h_{j,BMU(\vec{x})} = \exp\left(- \frac{\left\| \vec{w}_j - \vec{w}_{BMU(\vec{x})} \right\|^2}{2\sigma^2} \right) \qquad (3)$$

The size of the Gaussian neighborhood function is determined by parameter σ. In order to enforce a convergent behavior the size of neighborhood is reduced by decreasing the parameter σ. Typically, the exponential decay rule is applied. The learning rate η controls the rate of adaptation of individual neurons. Like the size of the neighborhood function, its value also exponentially decays with the elapsed training time.

The learning process described in steps 2-5 is repeated until a specific convergence criterion is met. This criterion is typically defined as the average weight change of all the neurons after every training cycle. Once the average weight change drops below a predefined value the training is terminated. An illustrative example of a 2D SOM in the input space adapted to a 2D distribution of data is shown in Fig. 1(b).

III. CAVE-SOM

A Cave Automatic Virtual Environment (CAVE) is an immersive virtual environment. CAVE uses multiple stereoscopic projections to project a 3D environment in a room sized cube [32]. It allows the user to immersively interact with a 3D environment using head tracking system and a 6 Degrees-Of-Freedom (DOF) motion tracking tools.

The CAVE itself consists of three walls and a floor with stereoscopic image projection (optionally a fifth projection wall can be added on top). The user is able to step inside the CAVE and interact with the virtual environment in a seamless way. 3D motion tracking is used to track the head movement of the user and the environment is updated accordingly. In addition, the user has a 6DOF interaction tool called the wand, which allows the user to further interact

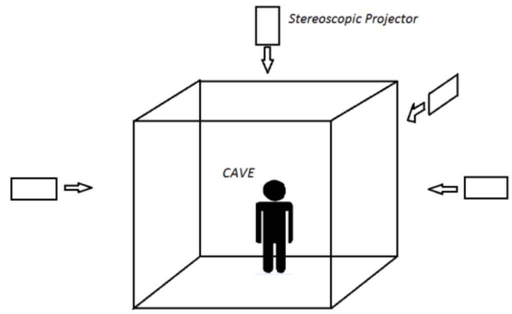


Fig. 2 The CAVE setup with 4 stereo-projection enabled walls.

with the CAVE environment (e.g. pointing, highlighting, zooming or menu selection). A schematic view of the CAVE system is depicted in Fig. 2.

The presented CAVE-SOM system utilizes the CAVE to visualize and interact with a 3D SOM structure. The system can display large multi-dimensional datasets in an immersive virtual environment that allows visual data mining using natural human perception. The CAVE-SOM combines powerful dimensionality reduction and feature extraction capability of the SOM algorithm with the immersive 3D visualization capability of CAVE. Fig. 3 shows two sample views of the developed system.

In addition, the application takes advantage of the available vision and motion tracking devices provided by CAVE to allow for interaction with the displayed datasets. The CAVE-SOM system thus allows real-time interactive visual discovery of clustering tendencies, differences between datasets or feature correlations. The system also provides a dynamically constructed Graphical User Interface (GUI) (see Fig. 4). This GUI is anchored directly in the 3D space. The software was implemented based on the VRUI library [33].

The implemented functionalities of the CAVE-SOM system can be further divided into several areas: SOM construction, neurons visualizations, connections visualizations and selection tools.

A. SOM Construction and Training

The architecture of the SOM can be dynamically created in the application. The number of neurons in each dimension can be specified.

In addition, the user has access to all control parameter of the learning process. Here, the initial size of the

Fig. 3 CAVE-SOM visual data-mining system.

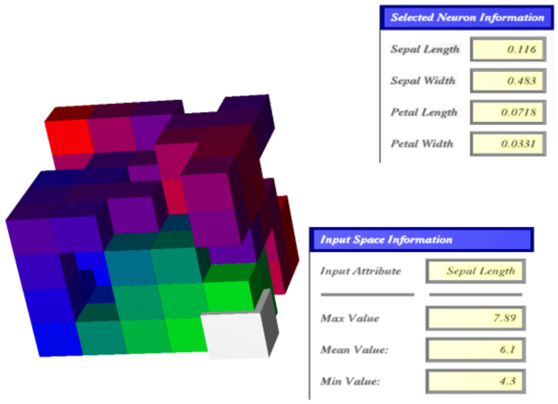


Fig. 4 Dynamic Graphical User Interface of CAVE-SOM.

neighborhood, the initial learning rate, and the rate of exponential decay of these parameters can be adjusted. The learning process of SOM is visualized in real-time.

B. Neurons Visualizations

The fundamental building unit of the 3D SOM structure is a neuron. Basic properties of 3D geometric neuron objects such as shape, size, color or transparency are utilized in the CAVE-SOM system to enhance the visual data mining process.

The CAVE-SOM system displays neurons as cubes. The position of each neuron is determined by its position in the 3D SOM grid. The size of each neuron can encode valuable information about the distribution of the input data throughout the SOM and about the relative importance of each neuron. In case of unlabeled data, the size of each neuron simply shows the frequency of patterns having particular neuron as their *BMU*. The size is thus proportional to the significance of the neurons. Neurons that were more frequently selected as *BMUs* are enlarged and they naturally attract more users' attention.

In case of labeled data, the shape and size of the neurons can be used to visualize distribution of patterns throughout the SOM for each specific class. In this mode, neurons that were assigned to more than one class are visualized as a composition of multiple scaled cubes with their size proportional to the frequency of patterns from each class (see Fig. 5).

Two additional modes have been implemented in the CAVE-SOM system to draw user's attention to specific features of the SOM: uniqueness and similarity of different classes for labeled datasets (see Fig. 10(b)). Here, the uniqueness and similarity are encoded using the size of neurons. The proposed uniqueness measure u_j of neuron j can be computed based on the histogram of classes' hits as follows:

$$u_j = \frac{\alpha}{C} \sum_{i=1}^{C} \left(\frac{H_j}{C} - h_j^i \right)^2 \qquad (4)$$

Here, H_j is the number of all hits for neuron j, h_j^i is the number of hits for class i for neuron j, C is the total number of classes and α is the normalization coefficient. The similarity s_j can be computed as its opposite value $s_j = 1 - u_j$.

Another elementary attribute of each visualized neuron is its color. The color is used in several visualization modes. Firstly, the distribution of values from a selected original input dimension can be projected onto the SOM (see Fig. 6 and Fig. 11). This view provides valuable information about the correlation of input dimensions and the distribution of input values in the output feature space. In case of labeled input data, the color can encode the provided output class (see Fig. 5).

The transparency of the displayed neurons proved to be very important when visualizing 3D SOM. In the CAVE-SOM system the transparency of neurons encodes the U-matrix. The U-matrix stores the information about the distance between neighboring neurons in the original input space (see Fig. 10(a)). The utilization of the advanced 3D visualization features of the CAVE system is especially important in case of transparency. Here, the CAVE offers improved depth-judgment and spatial relationship understanding in case of objects occlusion.

C. Connections Visualizations

The second important entity of the 3D SOM, which is displayed by the CAVE-SOM system, is the *connection strength* between neighboring neurons (see Fig. 8(a) and Fig. 9). The connections have been previously shown to encode valuable information about the input feature space [28], [29]. The CAVE-SOM system visualizes the connections as bars with variable width and transparency. The width and transparency encode the connection strength, which is proportional to the distance between the connected neurons in the input space. The scaling factor of the connection strength can be interactively adjusted, allowing visual discoveries of clustering tendencies in the 3D SOM. By visualizing the inverse of the connection strength the separation between neurons can be emphasized (see Fig. 8(b)). These separations can be viewed as cluster boundaries.

D. Selection Tool

The CAVE-SOM system also implements the *selection tool*, which allows the user to select and focus on a single neuron in the 3D SOM (see Fig. 7(b)). In this manner, the information about each neuron can be extracted. Furthermore, the rest of the SOM can be displayed with respect to the selected neuron, for instance color-encoding the mutual distances to each neuron. The users are then able to identify the topological neighbors of the selected neuron in the original input space.

IV. Experimental Results

The CAVE-SOM system was tested on two benchmark problems and applied to analysis of wind power data.

A. Benchmark Problem I – Iris dataset

The CAVE-SOM system was applied to analyze the well-known iris dataset [34]. The iris dataset is a common benchmark problem describing the separation among three

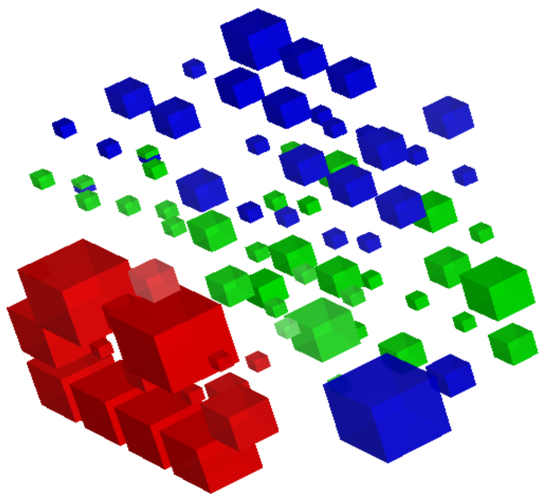
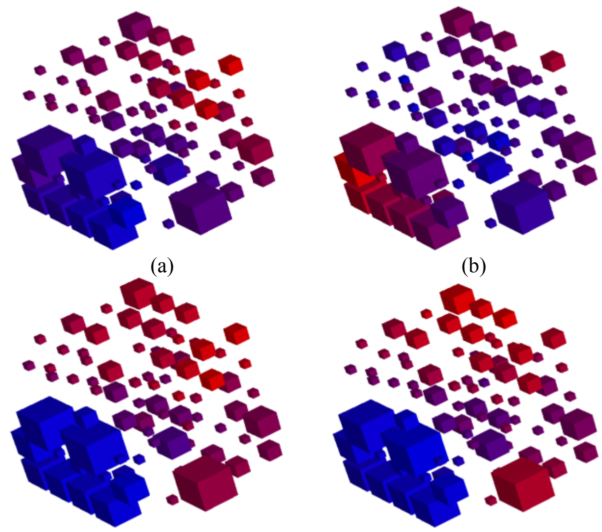


Fig. 5 Iris dataset in the CAVE-SOM – histogram view: Setosa (red), Virginica (green) and Versicolor (blue).



(a) (b) (c) (d)

Fig. 6 Iris dataset – input space view of sepal length (a), sepal width (b), petal length (c), and petal width (d).

species of Iris flowers – Setosa, Virginica, and Versicolor. Each data point is described using 4 features: the length and the width of the sepal and petal. The dataset consists of 150 patterns, divided into 50 patterns for each class.

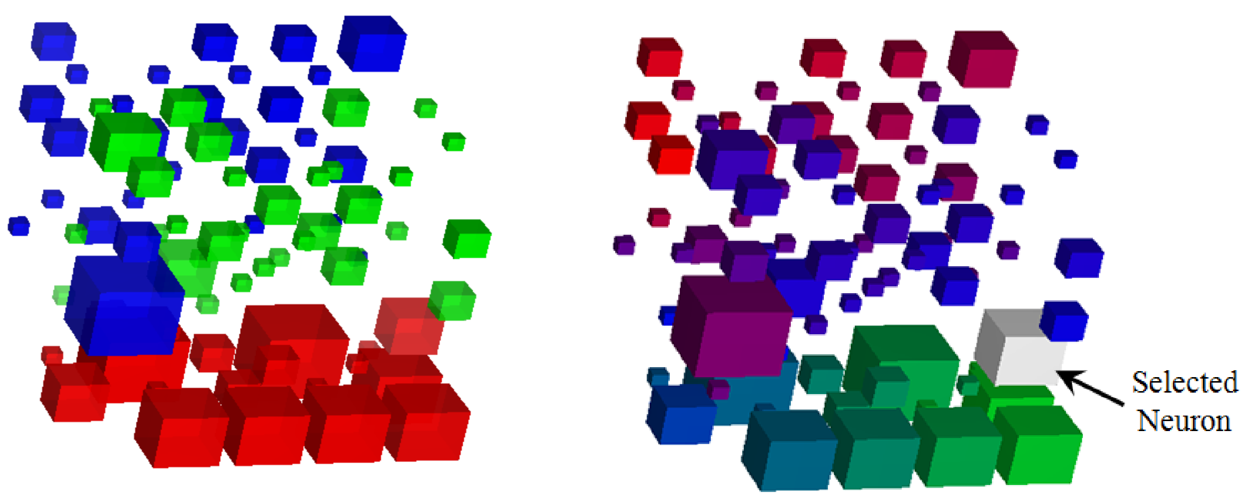
Fig. 5 shows the histogram view of the trained 3D SOM, with color-encoding of the classes as follows: red is Setosa, green is Virginica and blue is Versicolor. The neuron size encodes the histogram of the best matching units. The CAVE-SOM visually confirms the well known distribution of classes in the iris dataset. It can be observed that the Setosa class (red) forms a well-separated and compact cluster. On the other hand, the topological distribution of the neurons assigned to Virginica and Versicolor classes suggests linearly non-separable input data. In the matter of fact, several neurons in Fig. 5 can be seen to be assigned to both classes.

Fig. 6 shows the color-encoding of the input dimension values projected on top of the histogram view of the 3D SOM. Each figure shows the distribution of particular input dimension. By observing Fig. 6, the strong correlation between the petal length and the petal width attributes can be determined. In addition, by correlating the histogram and the inputs views in Fig. 5 and Fig. 6 the attributes of petal length and petal width can be seen to provide good separation between individual classes. These observations are in accordance with the expected results.

Furthermore, Fig. 7 demonstrates the use of the selection tool. The iris dataset is displayed using color encoding of classes in Fig. 7(a). However, this view lacks the information about the spatial closeness of neurons in the input space. Fig. 7(b) then shows the color encoding of the distances among neurons, when a particular neuron is selected. The green and red colors encode the smallest and the largest distance, respectively. The selected neuron belongs to the Setosa class, and thus the neurons belonging to Setosa class are encoded in green and blue due the small distance from the selected neuron. This view demonstrates the non-linear mapping performed by the 3D SOM.

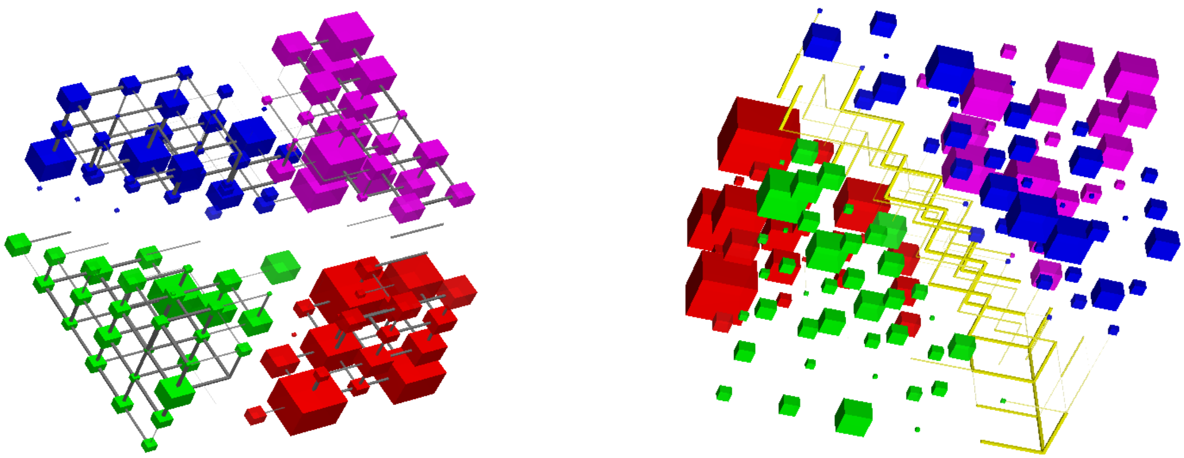
B. Benchmark Problem II – Gaussian Clusters

The second benchmark dataset used was an artificially generated dataset consisting of 10,000 data points generated in 10-dimensional input space. The data set contained four compact clusters with Gaussian distributions.



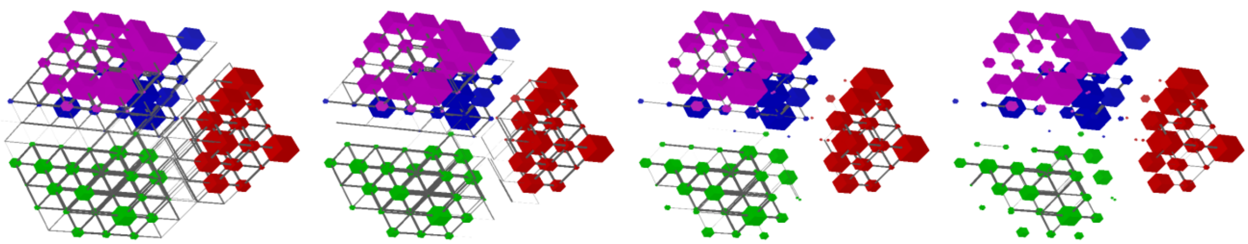
(a) (b)

Fig. 7 Iris dataset – histogram view (a) and neuron selection view (b). The selected neuron is highlighted in white. Green and red colors encode the smallest and the largest distance in the original input space, respectively.



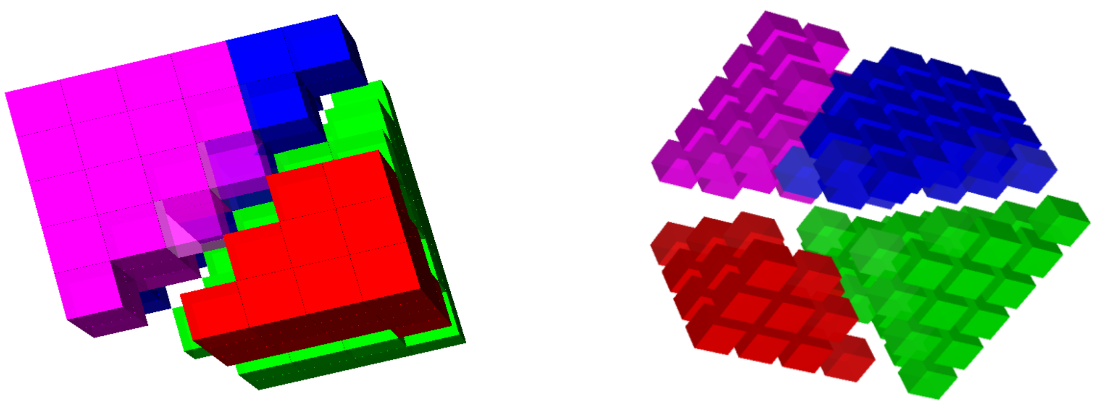
(a) (b)

Fig. 8 Gaussian dataset – connection view (a) and separation view (b).



(a) (b) (c) (d)

Fig. 9 Gaussian dataset – with decreasing scaling factor for connection strength (a)-(d).



(a) (b)

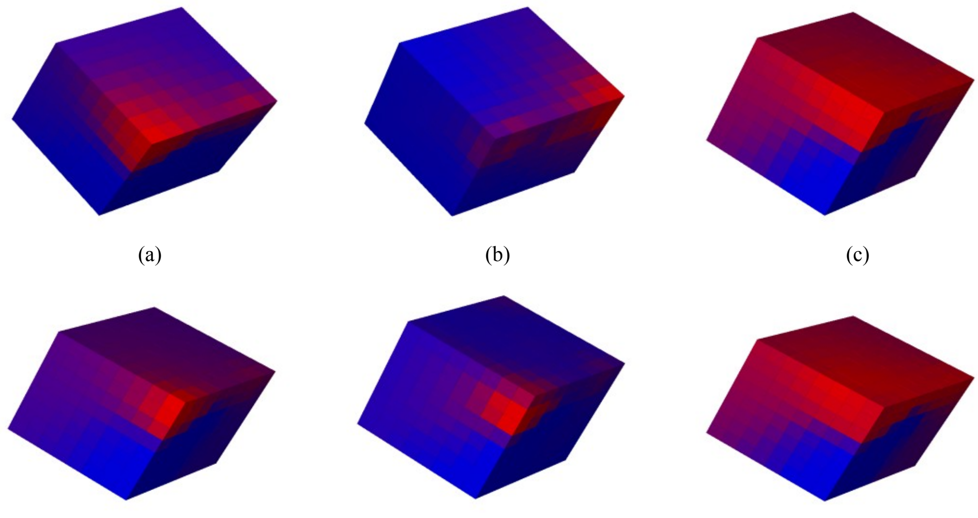
Fig. 10 Gaussian dataset – transparency encoding of U-matrix (a) and the uniqueness view (b).

Fig. 8 depicts the histogram view together with the color-encoding of classes in the trained 3D SOM. The visualization clearly shows the four compact clusters in the dataset. In addition, Fig. 8(a) and Fig.8(b) display the connections and the separations among the neurons, respectively. It can be seen that displayed connections further enforce the understanding of the clustering tendencies in the dataset. In order to display compact and well separated clusters the correct scaling factor for the connection strength must be set. This factor puts a threshold on the minimum connection strength to be displayed and further scales the width of the connection accordingly. The CAVE-SOM system allows the user to adjust this parameter interactively and updates the view in real-time allowing for intuitive and visual-based process of determining its optimal value. Examples of the histogram view with displayed connections for different values of the scaling factor are shown in Fig. 9.

Further, Fig. 10(a) shows the transparency encoding of the U-matrix. It can be observed that neurons located at the boundaries between classes disappear from the view, separating the 3D SOM into compact clusters. Fig. 10(b) then shows the size encoding of the uniqueness of each class. Again, this view provides good visual understanding of the clusters in the original multi-dimensional space.

C. *Wind-Power Data Analysis*

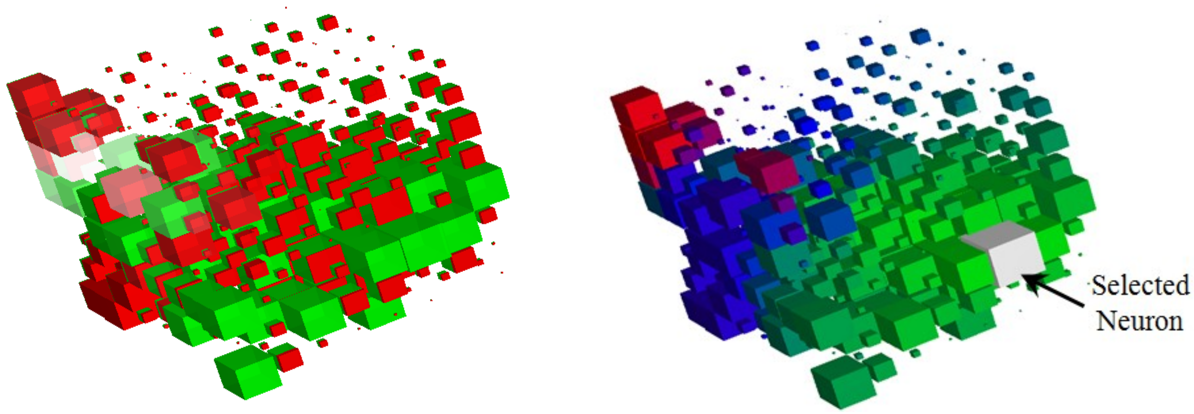
Finally, the CAVE-SOM system was used for analysis of wind-power generation dataset. Wind-power generation constitutes an active engineering area, which is becoming increasingly important with the recent emphases on



(a) (b) (c)

(d) (e) (f)

Fig. 11 Wind-power dataset - input view: generated power (a), accumulated daily energy production (b), turbine blades RPM (c), wind speed (d), standard deviation of the wind speed (e), and generated volts (f).



(a) (b)

Fig. 12 Wind-power dataset - uniqueness view (a), Idaho Falls (red), Gilpin (green), and neuron selection view (b).

renewable energy sources. The CAVE-SOM system was used to analyze the differences in wind-power production in two geographically distinct areas of USA, namely Idaho Falls, Idaho and Gilpin, Colorado.

The wind-power dataset was acquired through Wind Powering America program sponsored by the U.S. Department of Energy. This program is committed to increase the use of wind energy in the United States. The dataset consisted of the generated power, daily energy production, RPM of the turbine, wind speed and generated volts sampled every 30 seconds throughout June, 2010. The data were preprocessed using a sliding window technique, computing the average values of the attributes in the window. Also the standard deviation of the window speed was calculated. The preprocessed dataset then contained 30,000 6-dimensional data points.

Fig. 11 shows the input space view of the wind-power data projected onto the trained 3D SOM. An expected strong correlation between the RPM of the rotor blades and the generated volts in Fig. 11(c) and Fig. 11(f) can be observed. Similarly, the input dimension of generated power in Fig. 11(a) and wind speed Fig. 11(d) also feature high positive correlation. Using this visualization mode of the CAVE-SOM system, engineers are easily able to determine the correlation between different attributes of wind power production.

The goal of the performed analysis of the wind-power data is to obtain an insight into the differences between wind-power productions in geographically difference regions of United States. The results of such analysis can provide valuable information for determining the suitability of certain geographical area for further wind turbine construction. Fig. 12(a) shows the uniqueness view of the class color-encoding of the wind-power data. Here, the Idaho Falls and the Gilpin classes of data are highlighted in red and green, respectively. The uniqueness view reduces the size of neurons belonging equally to both classes and thus drawing user attention to the highlighted neurons that are unique for each class. Focusing on such neurons can provide valuable insight into the difference between the two classes. Using this visualization tool of the CAVE-SOM system users are able to determine the preferred location for a wind turbine based on the differences in terms of wind profile and desired power output. Selected unique behavior of interest can be highlighted for further analysis as demonstrated in Fig. 12(b).

Another application of the CAVE-SOM system for wind-power production analysis can be anomaly detection.

Training the SOM using known normal behavior of a wind turbine and then visualizing the uniqueness view of new available data in the SOM can aid fault diagnosis and wind turbine maintenance.

V. Conclusion

This paper presented the design and implementation of an immersive visual data-mining system – the CAVE-SOM. The 3D SOM algorithm was utilized to learn the spatial and topological relationships in the multi-dimensional data. The 3D SOM structure was then visualized in the immersive CAVE environment. The implemented tool allows users to explore and interact with the multi-dimensional data in a natural and intuitive way.

The implemented CAVE-SOM system was first validated on two benchmark problems, the iris dataset and a multi-dimensional Gaussian distribution. Next, CAVE-SOM was used for analysis of wind-power generation data. It was demonstrated that the algorithm successfully identifies data separations and clustering tendencies. Furthermore the system could be used analyzing important similarities and unique features of different data classes.

The knowledge extracted using the CAVE-SOM can be used for further informed decision making and machine learning.

References

[1] S. Sato, Y. Tateyama, "Super High Definition Three-Dimensional Display Environment Applied to Visual Data Mining," in *Proc. of the 13th International Conference on Network-Based Information Systems*, pp. 414-419, 2010.

[2] M. Bohlen, L. Bukauskas, P. S. Eriksen, S. L. Lauritzen, A. Mazeika, P. Musaeus, P. Mylov, "3D Visual Data Mining—Goals and Experiences," in *Computational Statistics & Data Analysis*, vol. 43, pp. 445-469, 2003.

[3] N. Sakr, F. A. Alsulaiman, J. J. Valdes. A. E. Saddik, N. D. Georganas, "Exploring the Underlying Structure of Haptic-based Handwritten Signatures using Visual Data Mining Techniques," in *Proc. of 2010 IEEE Haptics Symposium*, Waltham, Massachusetts, USA, pp. 467-474, March, 2010.

[4] J. J. Valdes, "Evolutionary Computation Based Nonlinear Transformations to Low Dimensional Spaces for Sensor Data Fusion and Visual Data Mining," in *Proc. of 2010 IEEE World Congress on Computational Intelligence – WCCI*, Barcelona, Spain, pp. 2242-2249, July, 2010.

[5] E. Mozzafari, A. Seffah, "From Visualization to Visual Mining: Application to Environmental Data," in *Proc. of the First International Conference on Advances in Computer-Human Interaction*, pp. 143-148, 2008.

[6] D. A. Keim, "Information Visualization and Visual Data Mining," in *IEEE Trans. on Visualization and Computer Graphics*, vol. 7, no. 1, pp. 100-107, January-March 2002.

[7] S. J. Simoff, M. H. Böhlen, A. Mazeika, *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics.* New York: Springer, 1998, pp. 1-5.

[8] E. J. Wegman, J. Symanzik, "Immersive Projection Technology for Visual Data Mining," in *Journal of Computational and Graphical Statistics*, vol. 11, no. 1, pp. 163-188, 2002.

[9] H. R. Nagel, E. Granum, P. Musaeus, "Methods for Visual Mining of Data in Virtual Reality," in *Proc. of the International Workshop on Visual Data Mining*, pp. 13-27, 2001.

[10] O. Zaiane, A. Ammoura, "On-line Analytical Processing While Immersed in a CAVE," in *Proc. of Second International Workshop on User Interfaces to Data Intensive Systems*, pp. 102-115, 2001.

[11] A. van Dam, D. H. Laidlaw, R. M. Simpson, "Experiments in Immersive Virtual Reality for Scientific Visualization," in *Computers & Graphics*, vol. 26, pp. 535–555. 2002.

[12] Xianjun Ni, "Research of Data Mining Based on Neural Networks," in *World Academy of Science, Engineering and Technology*, vol. 39, pp. 381-384, 2008.

[13] S. Okajima, Y. Okada, "Treecube+3D-ViSOM: Combinational Visualization Tool for Browsing 3D Multimedia Data," in *Proc. of The 11th International Conference Information Visualization*, pp 40-46, 2007.

[14] S. YU, H. LI, Q. Xu, X. Wu, "Fuzzy Self-organizing Maps for Data Mining with Incomplete Data Sets," in *Proc. of International Conference on Computer Application and System Modeling*, pp. 336-340, 2010.

[15] J. Vesanto, E. Alhoniemi, "Clustering of the Self-Organizing Map," in *IEEE Trans. on Neural Networks,* vol. 11, no. 3, pp. 586-600, 2000.

[16] T. Kohonen, *Self-Organizing Maps*. Series in Information Sciences, Vol. 30. Springer, Heidelberg. Second ed. 1997.

[17] Q. Ang, W. Wang, Z. Liu, K. Li, "Explored Research on Data Preprocessing and Mining Technology for Clinical Data Applications," in *Proc. of The 2nd IEEE International Conference on Information Management and Engineering*, pp.327-331, April 2010.

[18] J. Yan, J. Thill, "Visual Data Mining in Spatial Interaction Analysis with Self-Organizing Maps," in *Environment and Planning B: Planning and Design*, vol. 36, pp. 466-486, 2009.

[19] J. A. F. Costa, Y. Hujun, "Gradient-based SOM Clustering and Visualisation Methods," in *Proc. of The 2010 International Joint Conference on Neural Networks*, 2010.

[20] H. Jin, W. Shum, K. Leung, M. Wong, "Expanding Self-Organizing Map for Data Visualization and Cluster Analysis," in *Information Sciences*, vol. 163, no. 1-3, pp. 157-173, 2004.

[21] A. A. Miranda, Y. A. Le Borgne, G. Bontempi, "New Routes from Minimal Approximation Error to Principal Components," *Neural Processing Letters*, vol. 27, no. 3, June 2008.

[22] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers 18*, pp 401–409, 1969.

[23] J. B. Tenenbaum, V. de Silva, J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[24] S. T. Roweis, L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 22, pp. 2323–2326, December 2000.

[25] M. Gashler, D. Ventura, T. Martinez, "Iterative Non-linear Dimensionality Reduction with Manifold Sculpting," *Advances in Neural Information Processing Systems*, vol. 20, pp. 513-520, MIT Press, Cambridge, MA, 2008.

[26] M. A. Kraaijveld, "A Non-linear Projection Method Based on Kohonen's Topology Preserving Maps," in *Pattern Recognition*, pp. 41-45, 1992.

[27] J. Vesanto,"SOM-based Data Visualization Methods," in *Intelligent Data Analysis*, vol. 3, no. 2, pp. 111-126, 1999.

[28] T. Martinetz, K. Schulten, "Topology Representing Networks," in *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1993.

[29] K. Tasdemir, E. Merényi, "Exploiting Data Topology in Visualization and Clustering of Self-Organizing Maps," in *IEEE Trans.On Neural Networks*, vol. 20, no. 4, pp. 549-562, 2009.

[30] T. Kohonen, "Automatic Formation of Topological Maps of Patterns in a Self-Organizing System, " in *Proc. SCIA, E. Oja, O. Simula, Eds.* Helsinki, Finland, pp. 214-220, 1981.

[31] S. Haykin, *Neural Networks and Learning Machines – Third Edition*, Prentice Hall, 2008.

[32] E. Wegman, "Affordable Environments for 3D Collaborative Data Visualization," *Computing in Science & Engineering*, vol2, no 6, pp. 68-72, 2000.

[33] O. Kreylos, VRUI homepage [URL], Available: http://idav.ucdavis.edu/~okreylos/ResDev/Vrui/index.html, from January 2010.

[34] A. Asuncion, D. J. Newman, UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science