# Visualization of High-Density 3D Graphs Using Non-Linear Visual Space Transformations

Ming C. Hao, Umeshwar Dayal, Pankaj Garg, Vijay Machiraju
Software Technology Laboratory
HP Laboratories Palo Alto
HPL-2002-70
March 22nd , 2002*

E-mail: (ming_hao, dayal, garg, vijay_machiraju) @hp.com

visualization, non-linear 3D space transformation, high-density

The real world data distribution is seldom uniform. Clutter and sparsity commonly occur in visualization. Often, clutter results in overplotting, in which certain data items are not visible because other data items occlude them. Sparsity results in the inefficient use of the available display space. Common mechanisms to overcome this include reducing the amount of information displayed or using multiple representations with a varying amount of detail. This paper describes our experiments on "Non-Linear Visual Space Transformations" (NLVST). NLVST encompasses several innovative techniques: (1) employing a histogram for calculating the density of data distribution; (2) mapping the raw data values to a non-linear scale for stretching a high-density area; (3) tightening the sparse area to save the display space; (4) employing different color ranges of values on a non-linear scale according to the local density. We have applied NLVST to several web applications: market basket analysis, transactions observation, and IT search behavior analysis.

* Internal Accession Date Only Approved for External Publication

Published in and presented at the SPIE Conference on Visualization and Data Analysis, 20-25 January 2002, San Jose, CA

# Visualization of High-Density 3D Graphs Using Non-Linear Visual Space Transformations

Ming C. Hao, Umeshwar Dayal, Pankaj Garg, Vijay Machiraju
Hewlett Packard Research Laboratories, Palo Alto, CA

## ABSTRACT

The real world data distribution is seldom uniform. Clutter and sparsity commonly occur in visualization. Often, clutter results in overplotting, in which certain data items are not visible because other data items occlude them. Sparsity results in the inefficient use of the available display space. Common mechanisms to overcome this include reducing the amount of information displayed or using multiple representations with a varying amount of detail. This paper describes our experiments on "Non-Linear Visual Space Transformations" (NLVST). NLVST encompasses several innovative techniques: (1) employing a histogram for calculating the density of data distribution; (2) mapping the raw data values to a non-linear scale for stretching a high-density area; (3) tightening the sparse area to save the display space; (4) employing different color ranges of values on a non-linear scale according to the local density. We have applied NLVST to several web applications: market basket analysis, transactions observation, and IT search behavior analysis.

**Keywords:** Visualization, Non-Linear Visual Space Transformation, High Density, Stretch, Distance, Color.

## 1. INTRODUCTION

With real world data, we discover that the distribution is seldom uniform. As a consequence, clutter and sparsity commonly occur in visualization. Often, clutter results in overplotting, i.e., certain data items are not visible and the overall structure gets lost. Sparsity, on the other hand, results in inefficient use of the available display space. Figure 1A illustrates that the graph generated from a conventional relation-based visualization [4,5,6] has a high-density area (circled with red). Data items are too close to each other and penetrate each other. The visualization is very congested and hard to navigate. In addition, data similarities are not fully reflected by the computed colored values. Figure 2A illustrates that pseudo coloring schemes fail when all the similarity edges are colored green due to the non-uniform dataset.

Common mechanisms to overcome the above problems are either to reduce the amount of information displayed or to use multiple representations with a varying amount of detail. Besides, graph visualization methods that deal with the non-uniform distribution of data [1, 2, 8, 11] have been very popular in information visualization. For instance, Woodruff's "Visual Information Density Adjuster" [3] helps users construct applications in which the overall display density remains constant. Frank [9] used multiple representations for cartographic objects in a multi-scale tree. Recently, using the density of visual items, Peter Pirolli and Stuart K. Card [10] from Xerox PARC integrated a theory of visual attention with information foraging theory in a focus + context visualization.

For instance, it is difficult to visually analyze web customer's purchasing behavior and response time from millions of web transactions. As the volume of transaction data grows, web analysts want to visually analyze information to make various decisions. To date, the visualization of large, dense, and non-uniform datasets has been required by many web applications, such as market basket analysis, web access observation, and IT resource management. The following are their requirements for a new visualization system to layout useful information:

(1) Stretch the high-density areas to prevent visual clustering and overlay.
(2) Employ different levels of color scales to differentiate closely similar data items.
(3) Scale to large volumes of data.

Correspondence: Email: (ming_hao, dayal, garg, vijay_machiraju) @hp.com

(A) Before Stretching (B) After Stretching

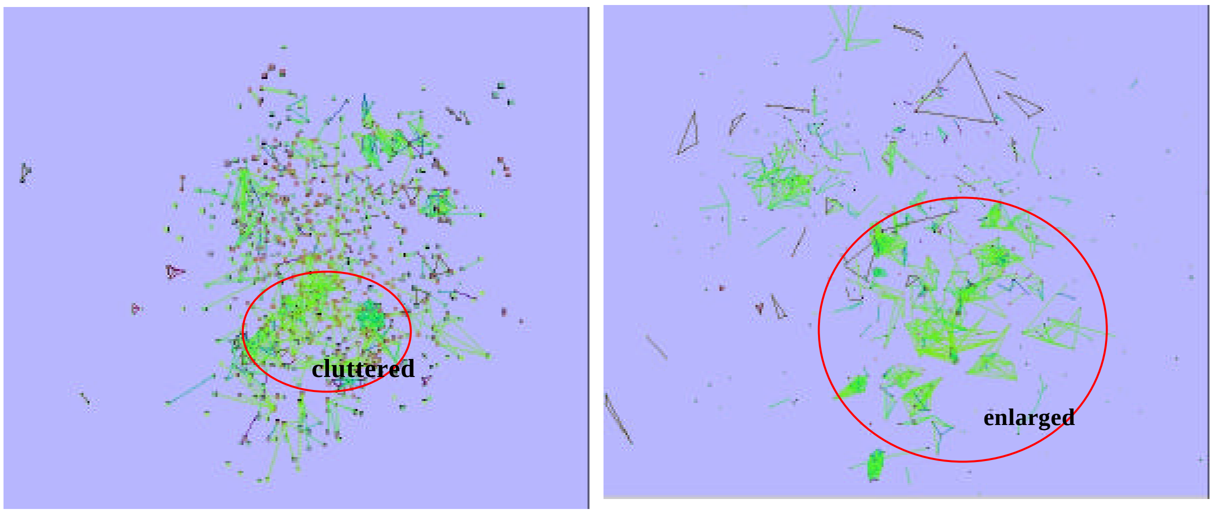



**Figure 1: Stretching A High-Density Area**
**(HP Shopping: # of Transactions: 355,776; # of Products: 959)**

(A) Before non-linear transform (all green) (B) After non-linear transform (multi-color)

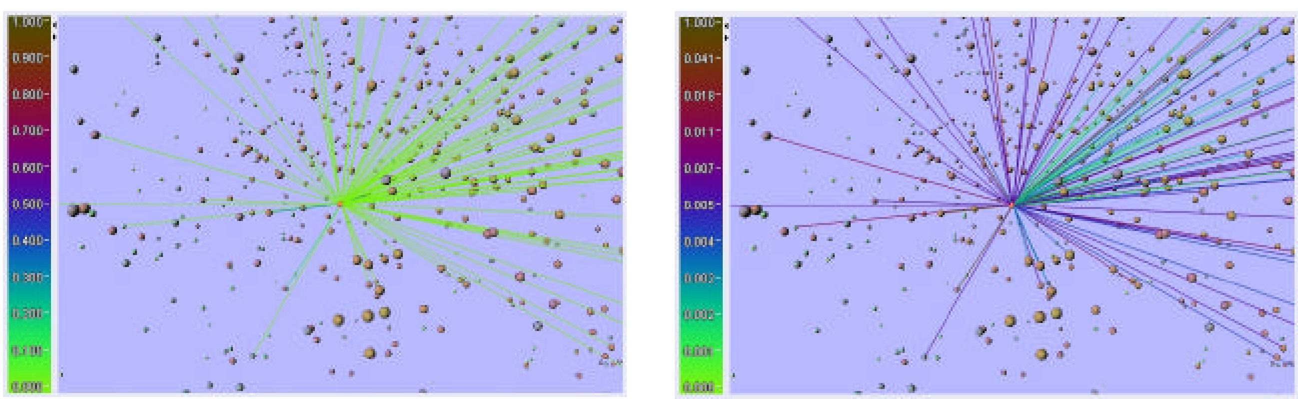


**Figure 2: Local Scale Transformation (shows low range of similarity values)**

## 2. OUR APPROACH

To date, many practical research projects have shown the usefulness of a physics-based mass-spring engine [4, 5, 6] for visualizing large web transaction data sets. NLVST is built on this engine to layout related data items onto a 3D graph. The relationships between data items are computed from the similarity. The strength of the relationships transforms to the *distance* between them. Data items with high relationships are placed close together.

At HP Laboratories, we have applied "Non-Linear Visual Space Transformations" (NLVST) to the high-density data sets to prevent cluttering and overlapping. Figure 1B shows a 3D graph after visual space stretching. NLVST tightens the sparse area to save the display space. The transformed graphs are consistent with the un-stretched graph; the relationships among data objects remain unchanged, but visually, the graph is not cluttered and easy to navigate.

NLVST employs colors to represent the similarity values. For non-uniform distributed data with a wide range of values, NLVST employs a local scaling transformation. Some data distributions exhibit large dynamics with few extremely high values and many extremely low values. It is hard to visualize each of these values on a linear scale. NLVST transforms the low range values on a non-linear scale according to local density. After these transformations, the similar-value data items can be identified and visualized, as illustrated in Figure 2B.

## 3. NON-LINEAR VISUAL SPACE TRANSFORMATION PROCESS

The non-linear visual space transformation process contains two key elements: distance and color. Distance represents the relationship between each pair of items. Color represents the similarity value between each pair of items. A dense graph often has items with very short distances between items (Figure 1A) and non-distinguished colored lines (Figure 2A, almost all green).

NLVST is built on Java-based multi-threaded parallelism. The NLVST process contains three basic components:

- Density analysis
- Distance transformation (stretch/tighten)
- Color transformation (multi-level)

Figure 3 illustrates the NLSVT overall process flow. Each of the above components is described further in the following sections.

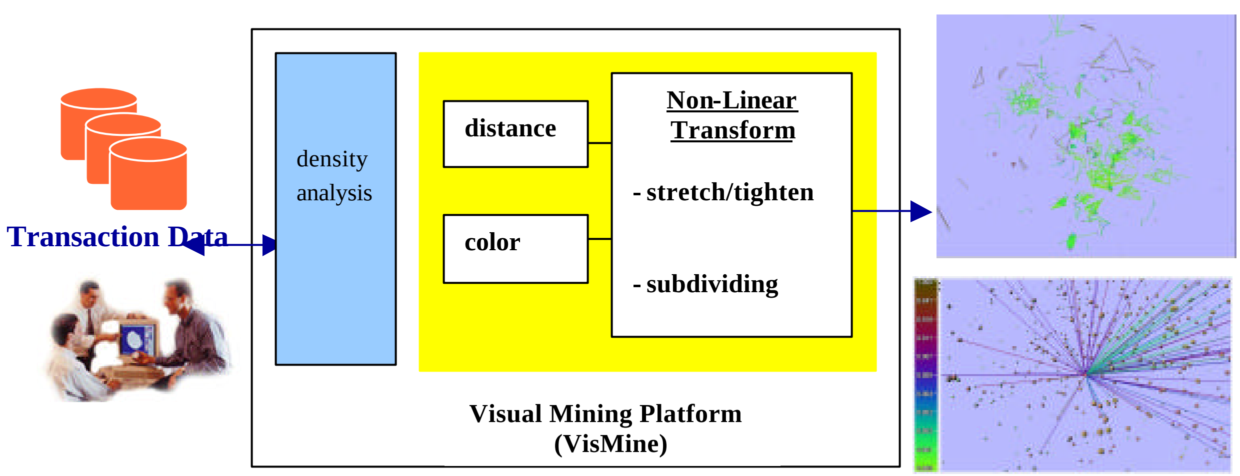



**Figure 3: Our Approach**

### 3.1. Density Analysis

NLVST employs a histogram to represent the density of the similarities among all the related items extracted from the web transaction data. As illustrated in Figure 4, the x-axis represents the relationships between items in the range of [0,1]. For example, in a market basket analysis, the tightness of the relationship is computed from the frequency of purchasing products (items) together. The y-axis represents the number of lines (edges) connecting pairs of data items. There are in the range of 0 to 6,994 edges. The data items are not uniformly distributed in the graph. The density of the graph on the left side in Figure 4 is very high. Most of the items are cluttered together in an area of the similarities close to 2.2%.

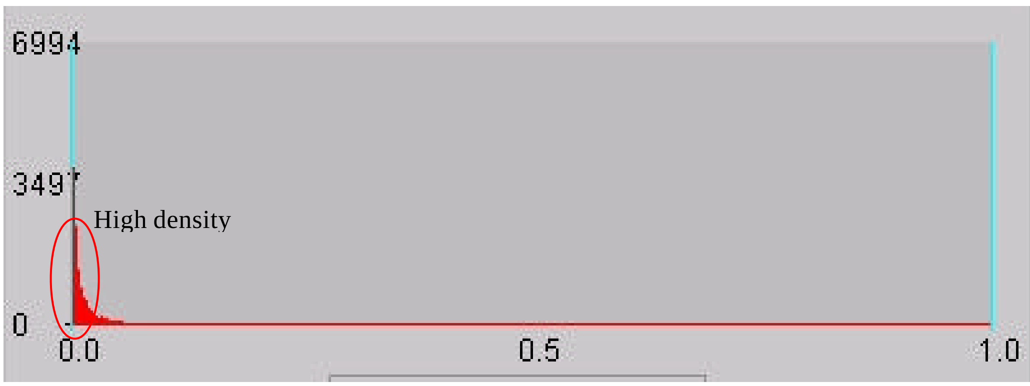



Figure 4: Density Histogram (HP Shopping: 355,776 Transactions, 959 Products)

### 3.2. Visual Space Transformations

There are two types of non-linear visual space transformations performed in NLVST: distance and color. First, NLVST transforms the distance between data items. Using non-linear transformation algorithms, NLVST computes the distance between data items. In a high-density area, NLVST automatically stretches the distance between tight data items. In a sparse area, NLVST shortens the distance between items with similarity values close to zero.

Second, NLVST computes the local density of a selected data item. NLVST transforms the color range according to the data local density. NLVST employs different color scales to represent the relationships between related data items. The bright color (e.g., green) represents weak relationships. The dark color (e.g., brown) represents strong relationships. Often, the color range can be very large, such as from 0 to 10,000,000. NLVST subdivides this large range into many different levels of scales.

For transformation consistency, both transformations are processed using a uniform non-linear scale. The detail processes are described in the following sections.

#### 3.2.1. Distance Transformation

From the density histogram, NLVST divides the visual screen into n density regions. For high-density regions, NLVST uses a non-linear visual space transformation algorithm to map original linear scale similarity values (relationship tightness) to a non-linear scale, such as log, before clustering the data items in a 3D space. The local log scaling function is derived from the ratio of the current data density over the range of the minimum and maximum tightness of the relationships. To avoid the graph reaching infinity, a small s will be used as an adjust factor. After transformation, the cluttered area will be mapped to a stretched scale; and the sparse area will be reduced to a shrunken scale.

### 3.2.2 Color Transformation

The general methods for a local scale transformation are: (a) to find the local density; (b) to transform the original data similarity to a ratio computed from the following local transformation scale function F; and (c) to impose a new non-linear subdivision on the scale with respect to the entire graph.

The local transformation scale function F is:

F = [# of edges with similarity <=similarity of the current data item] / total # of edges in a graph

## 4. APPLICATIONS

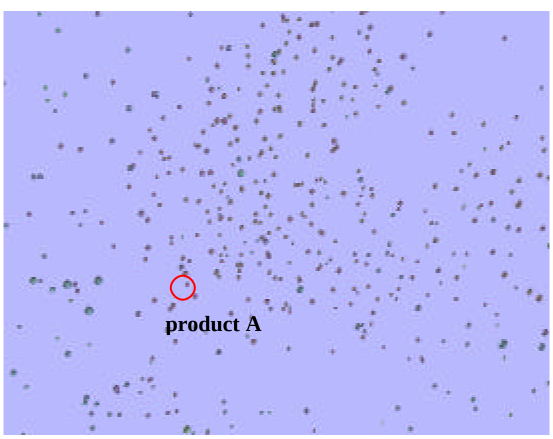
There are many web applications that can employ the non-linear visual space transformations for visualizing high-density non-uniform distributed 3D graphs. In Hewlett Packard Research Laboratories, we have applied NLVST to two web applications: market basket analysis and web access observation.

### 4.1. Market Basket Analysis

Figure 5A illustrates the use of the NLVST visualization technique for market basket analysis of real data taken from a Hewlett-Packard shopping web site. The vertex (sphere) represents products. The distance between products represents the frequency that product items are bought together. From 356,000 web transactions, there are about 1,000 different products (represented as cubes) and 21,100 active edges (lines). The color of the edge is used to show the closeness of similarities.

After performing non-linear visual space transformations, business analysts are able to navigate high-density 3D graphs generated from real data and to answer questions as to which product items are frequently bought together and how strong their correlations are. During the analysis, business analysts can click on a product to find its relationship with other products connected with different colored lines, as illustrated in Figure 5B.

**(A) A High Density Graph (products are not too close to each other)**




**(B) A New Non-linear Sub-division Scale (colored lines represent different correlations)**

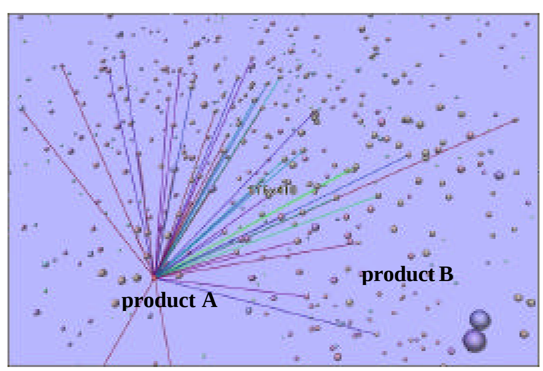



**Figure 5: Stretched Web Market Basket Analysis Visualization**

### 4.2 Web Access Observation

The second example (Figure 6) shown is the Hewlett Packard web transaction observation application. One common problem web analysts want to solve is how to use web transaction history to discover the clients causing network bottlenecks. The NLVST technique has been used experimentally to visually analyze web client behavior with respect to response time.

Figure 6 illustrates a graph generated from a web transaction observation data set. It contains 35,000 transaction records. There are 986 clients with over two thousands URLs. The rectangles represent clients that make transactions on the web. The spheres represent URLs. NLVST places clients with similar response times near to each other without penetrate each other and with their accessed URLs placed around them. The web analysts can easily click on a client and find their accessed URLs.

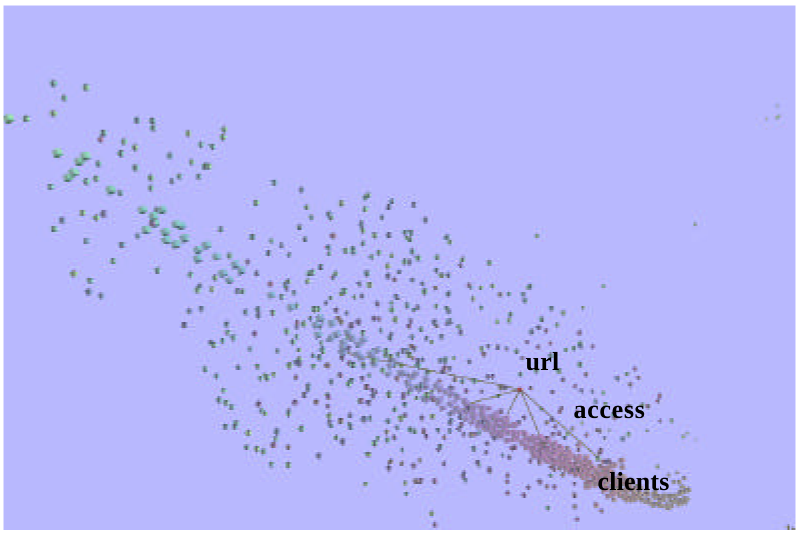



**Figure 6: Web Access Observation Visualization**

## 6. CONCLUSION

To address the requirements of high-density 3D graph visualization, we integrated a non-linear algorithm with distance and color transformations. This coupling enables a user to be able to visually analyze large volumes of transactions without cluttering the display. Future work will concentrate on the development of visual analysis interactions on density, such as multi-level zoom and pin techniques.

## 7. ACKNOWLEDGEMENT

Thanks to Sharon Beach, Aad van Morrsel, and Klaus Wurster from HP Software Technology Research Laboratory for their encouragement; thanks to Prof. Markus Gross and Daniel Cotting from the Swiss Federal Institute of Technology, Zurich, Switzerland and to Prof. Daniel Keim from the University of Constance, Germany for their suggestions and comments.

## 8. REFERENCES


1. Y. Ioannidis *et al*, “User-Oriented Visual Layout at Multiple Granularities”, Proc. AVI’96, Gubbio, Italy, May 1996.
2. E. R. Tufte, The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT. 1983.
3. A. Woodruff, J. Landay, M. Stonebraker, “Constant Information Density in Zoomable Interfaces“, Proc AVI’98, L’Aquila, Italy, May 1998.
4. T. C. Sprenger, R. Brunella, M. H. Gross, “H-BLOB: A Hierarchical Visual Clustering Method Using Implicit Surfaces”, IEEE/VIS2000.
5. T. C. Sprenger, M. H. Gross, “Ivory – An Object Oriented Framework for Physics-Based Information Visualization in Java”, IEEE InfoVis98, North Carolina.
6. M. H. Gross, T. C. Sprenger, J. Finger: “Visualizing Information on a Sphere”, Arizona, IEEE VisInfo97.
7. Ming Hao, Umesh Dayal, Meichun Hsu *et al*, “A Java-based Visual Mining Infrastructure and Applications”, IEEE InfoVis99, CA.
8. Yannis Ioannidis, Miron Livny, Jian Bao, Eben M. Haber, “User-Oriented Visual Layout at Multiple Granularities”, Silicon Graphics.
9. Andrew U. Frank, Sabine Timpf, “Multiple Representations for Cartographic Objects in a Mulit-Scale Tree – And Intelligent Graphical Zoom”, Computer & Graphics, Vol.18, No. 6, 1994.
10. Peter Pirolli, Stuart K. Card, Mija M. Van Der Wege, “Visual Foraging in a Focus + Context Visualization”, SIGCHI’01, March 31, Seattle, WA.
11. M. B. Carmo, J. D. Cunha, “Visualization of Large Volumes of Information Using Different Representations”, InfoVis’97, 1997, Los Alamitos, CA.