Visual Mining of E-Customer Behavior Using Pixel Bar Charts

Ming C. Hao, Julian Ladisch*, Umeshwar Dayal, Meichun Hsu, Adrian Krug
Hewlett Packard Research Laboratories, Palo Alto, CA.
(ming_hao, dayal)@hpl.hp.com; akrug@hpugrca.grc.hp.com

Abstract

To date, one of the common problems e-business and e-service want to solve is how to use their large volumes of sales histories, web transactions, and information resource data to know their customers' behavior. They want to understand their customers to develop long-term relationships and to improve service quality. Many data analysis methods use bar charts which only show highly aggregated data and lose most of the detail information due to data aggregation. This paper discusses the integration of a recent pixel bar chart technique that does not lose information into a visual data mining system. Our experiments show that the system allows the discovery of a wide range of patterns and trends. For example, it helps business analysts to identify the most profitable customers, the most frequent search key words, and the best sales timeline.

1. Introduction

Recently, the rapid increase of transactions on the Internet has led to the availability of large volumes of customer data. Business research efforts [2, 3] have focused on how to turn raw data into valuable information. For example, by exploring customer data, the business analysts are able to find and retain their most profitable customers and evolve their business strategies.

To date, a number of visualization applications have shown the usefulness of pixel-oriented techniques [1, 6, 7, 8, 9, 10] for analyzing millions of data records. Individual data items, such as customers, can be represented by pixels on a screen. The number of customers can as large as 1080*1024 (over 1.1 million) pixels. For example, the VisDB [1] system uses each pixel to represent one data value. Each pixel is arranged and colored to indicate the item's relevance to a user query. VisDB allows the user to explore multidimensional databases. Recently, pixel-oriented techniques have been used to build interactive decision tree classifiers for a multidimensional visualization of training data [7,8].

To analyze large volumes of web datasets, a common method is to use bar charts. However, regular bar charts require a high degree of data aggregation. Usually, valuable information gets lost. Based on our practical usage and design experience, we have discovered that regular bar charts are too restrictive. As illustrated in Figure 1A, regular bar charts usually only show 10 to 100 (aggregated) data values. The regular bar chart loses a large portion of the screen due to the large bars of different heights.

2. Visual Mining the Behavior Data

Customer behavior often involves relationships that need to be linked to different attributes of a dataset. For example, the purchase amount is related to the timeline, number of visits, and quantity. The customer search criteria are related to the search types and number of keywords.

This paper discusses the integration of a newly developed pixel bar chart technique [4] into a visual data mining system [5]. Pixel bar charts generalize regular bar charts but do not require data aggregation. They combine the basic idea of X-Y Diagrams with bar charts to allow an overlap free, non-aggregated display of large amounts of multi-attribute data. The color of each pixel represents an attribute value of a customer, e.g., search criteria, purchase amount, or number of keywords. The detailed information on each data item can be displayed by drilling down as needed.

In Figure 1A, the bar chart shows only the total purchasing dollar by month. No other related information is shown. In Figure 1B, the pixel bar chart shows the customer purchasing activities in one year by month. Analysts can easily discover the most profitable customers, the number of visits, and other purchase patterns, such as that the most profitable purchases occurred during the early months (the top area of the bars for months 2,3,4,5 have the most purple color).

In addition, as illustrated in Figure 1B, the pixel bar chart uses all the available screen space to cluster related data together. The display will not be cluttered as the data items increase in number.

*Presently with Computer Science Institute, Martin-Luther-University, Halle-Wittenberg, Germany
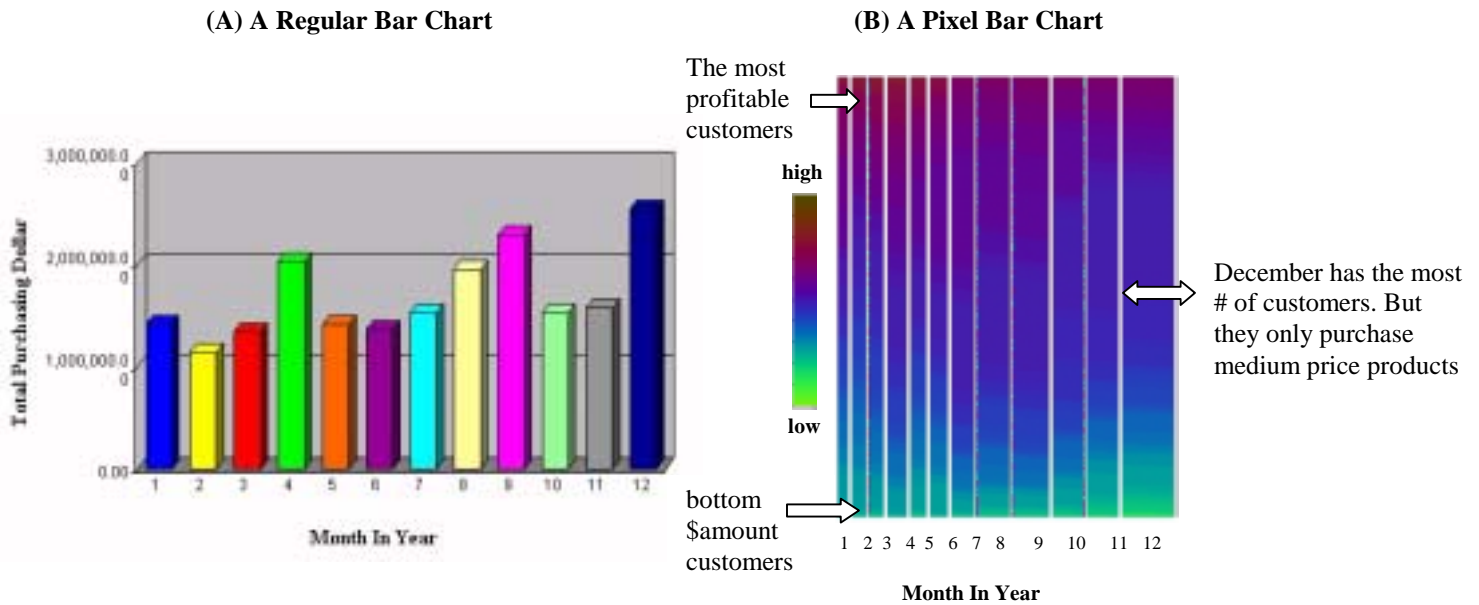julian@ladisch.de

(A) A Regular Bar Chart   **(B) A Pixel Bar Chart**

Figure 1: An Example of Mining 44,401 Sales Transactions By Months In A Year

3. Construction of Pixel Bar Charts

A general pixel bar chart integrates the idea of bar charts with X-Y-diagrams. It is ordered in the X-Y directions according to two attributes such as months in a year and purchase $amount. In general, a pixel bar chart can be specified as a five tuples:

<pixel object,
dividing attribute,
Y-ordering attribute,
X-ordering attribute,
coloring attribute>

For example, the five tuple used for generating the pixel bar chart shown in Figure 1B is:

    <customer,
    month,
    dollar amount ,
    number of visits,
    dollar amount >

Note that the X-ordering within each bar according to the number of visits is not visible in Figure 1B but will be visible in the multi-pixel bar chart shown later.

3.1 A Basic Pixel Bar Chart

Constructing pixel bar charts consists of: (1) dividing the x-axis space by grouping the pixels into rectangles according to the grouping attribute (e.g., months); (2) filling the rectangles with pixels from the bottom and placing them in order inside each rectangle according to the pixel ordering attribute (e.g., dollar amount for Y-ordering and number of visits for the X-ordering); (3) coloring the pixels according to the pixel coloring attribute (e.g., dollar amount in Figure 1B or the number of visits in Figure 2).
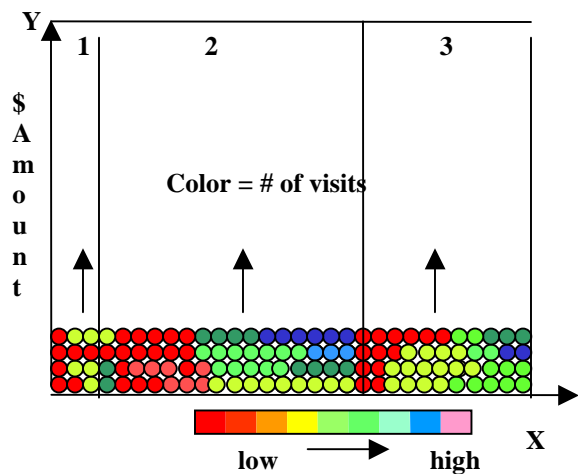

Figure 2: A Pixel Bar Chart Construction

3.2 Multi-Pixel Bar Charts

Multiple bar charts use the same arrangement of pixels but different attributes such as number of visits, quantities, and locations are mapped to

2

different colors (potentially with a non-linear scaling).

Multi-pixel bar chart are linked. Each pixel (i.e., customer) resides at the same relative location across all pixel bar charts but shows different attributes, such as number of visits, quantities, and locations… The color of each pixel varies based on the value of the corresponding attribute. The user can click on a pixel to get the customer's corresponding attribute values.

4. A Visual Data Mining System

To analyze large volumes of multi-attribute e-customer data, we have integrated the pixel bar chart technique into a data mining visualization system. The system places similar customers close to each other on the display. The "location" of the customer pixels in the pixel bar chart represents the similarity of their behavior.

The system uses a web browser with a Java activator. It provides a real-time, web-based interactive data mining and visualization environment. The analyst can access the data warehouse and visually mine the data in an integrated way.

4.1 Component Architecture

As illustrated in Figure 3, the e-customer mining system follows a three step process:
1. Placement
   The pixel bar chart is partitioned based on one or two customer attributes: e.g., timeline and search type. One pixel represents one data item, i.e., one customer. The ordering of pixels (y-axis) is based on attribute values, e.g., purchase amount or search type.

   The grouping algorithm consists of the sorting and pixel-filling mechanisms. The maximum and minimum values for each attribute are consistent across all groups.

2. Coloring
   The system uses a range of distinct colors to link multiple attributes. Color is calculated from the value of a selected attribute (such as purchase $amount, number of search keywords). For correlation, the location of each data item remains the same across multiple bar charts.

3. Exploration

The system provides mechanisms for simultaneous browsing and navigating among multiple attributes. The user is allowed to select, link, and retrieve data from the warehouse as needed.
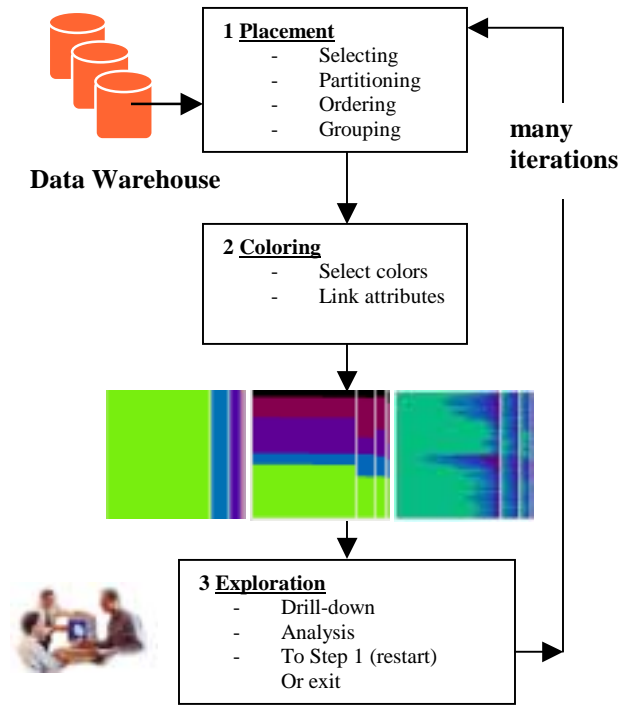

Figure 3: An Overview of E-Customer Visual Mining System

The above three process steps can be repeated until the user discovers the desired information.

4.2 Interaction and Pattern Discovery

Interactivity is an important aspect of the pixel bar chart system. Figure 4 shows the user interaction window of the system. In Figure 4, the user has constructed a pixel bar chart showing the customer purchasing activities of 12 months. The dividing attribute (x-axis) is "Month". The ordering (vertical and horizontal) attributes are "price" and "number of orders". The coloring attribute is "quantity". The attributes used for dividing, ordering, and coloring attributes can be selected and changed at execution time. There are four types of pull down menus for the user to select input data and to construct pixel bar charts.

To mine large volumes of multi-attribute data, the user may want to try many different data

arrangements. The system provides a "Recalc" button for the user to re-calculate, re-select, re-group, and re-visualize the pixel bar charts. The detail information is displayed in the right lower corner of the window.

Pixels reside at the same location across multiple pixel bar charts. The users can easily select a pixel to find the customer related information, i.e., number of visits, quantities, and locations.
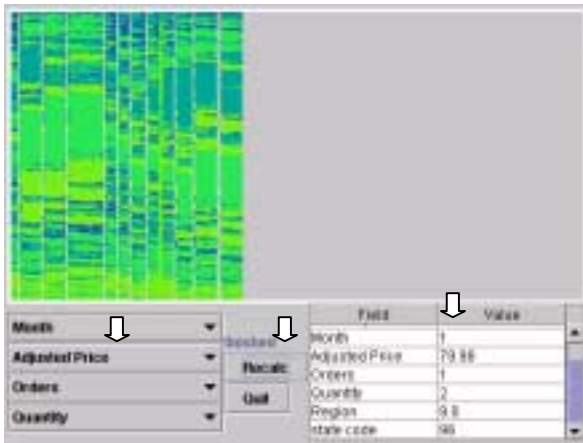

Figure 4: The User Interaction Window

5. E-Customer Behavior Experiments

Our e-customer behavior experiments consist of: (1) analyzing searching behavior; (2) finding shopping behavior.

5.1 Searching Behavior Mining

We applied pixel bar chart techniques to customer search behavior on HP's electronic support site (http://itrc.hp.com). As illustrated in Figure 5, each pixel represents one customer's search transactions. We mapped the used search criteria (Boolean, AllWords, AnyWord, Phrase, labeled 1-4 in Figure 5) together with the selected search type (Product Search, Solve/Fix Problem, Patch Search…) and the number of used keywords (i.e., printer, patch…) to generate the pixel bar chart. This technique places customers with similar searching behaviors next to each other based on the above-described parameters. Using this technique we were able to visualize the log entries from one month (several 100 thousands of search record entries) into one consolidated display. The visualization allowed us to detect certain customer behaviors as described in detail in the following paragraphs.

In our current search engine interface implementation, our customers use an average of 1.7 keywords per search. This is not enough to return significant relevant results. Based on our current research we enlarged the search engine query box to increase the number of used keywords in the query. But the average number of used keywords per search was not raised significantly. Using pixel bar charts we could easily identify a set of searches with a much higher average number of keywords (nearly 6).

In further analyzing the visualization we found a customer search cluster for the search type "Fix/Solve a problem" (marked area in Fig. 5C). Based on this data we can derive that the number of used keywords is also dependent on the search type used. In the next product releases, this information will be used to enhance the UI to make the usage of more keywords in corresponding searches easier.

In addition, we discovered that most of the searches had been applied without changing the search criteria value settings on the screen (the big green area in Figure. 5A). The defaults should represent the best choice for the customers' specific searches.

Most searches for a patch consisted of one keyword string, which probably represented the patch id number and not a real query string (marked area in Figure 5C).

Currently we are using the pixel bar charts to analyze and compare the customer search behavior obtained from different HP search engines and customer segments. We expect to get a better understanding of important design issues to be able to further improve the interface.

The pixel bar chart was the first visual data mining technique that allowed us to visualize huge sets of data in a limited space with a good overview and still retained the capability to drill down into details, because each pixel represents an individual search record.
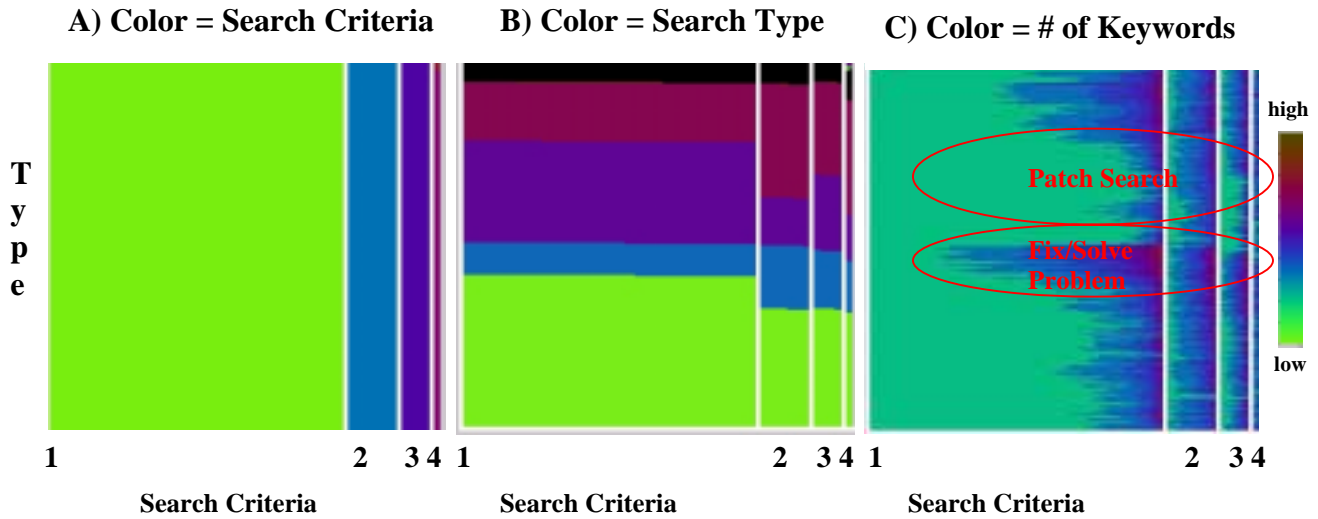
A) Color = Search Criteria   B) Color = Search Type   C) Color = # of Keywords

Figure 5: Pixel Bar Charts for Mining 106,199 IT Resource Center Customer Activities

5.2 Shopping Behavior Mining

To face today's business challenges, the company analysts want to apply e-customer purchase behavior to product selling and promotion. They want to know which month has the most sales, and who are their most profitable customers.

Figure 6 shows how to use a multi-pixel bar chart to explore the data describing customers' behaviors and understand their needs. With the pixel bar chart techniques, the business analyst transfers massive amounts of raw sales data and converts it into visualizations which help to understand customer behavior. This helps companies to strategize their campaigns.

The four pixel bar charts of Figure 6 are constructed as follows:

- *Time type* is the dividing attribute on the x-axis (12 months).
- *Purchase Dollar amount* is the y-ordering attribute.
- *Month, Purchase dollar amount, number of visits,* and *quantity* are the four coloring attributes.

Many important observations may be discovered in Figures 6 A-D. In the bars for the different attributes, the analyst may observe the following:

a) Time attribute
The 12 vertical bars with equal height represent 12 different months. (labeled 1-12 in Figure 6 A). The width of the bar indicates the number of customers. Month 12 (the largest area) is found to have the largest number of customers. Month 2 (the smallest area) has the least number of customers.

b) Purchase dollar amount attribute
Months 2, 3, 4 and 5 have the most top dollar amount customers (purple red). The dollar sale amount of month 12 is in the medium price range although it has the most number of customers (more green and less blue). It is interesting that especially customers with high dollar amounts tend to come back more often.

c) Number of visits attribute
The beige and green color distribution in the months of 3, 4, 5, and 6, indicate that customers for these months come back more often than customers for the other months. Christmas customers are mostly one-time customers.

d) Quantity attribute
The yellow and green colors spread largely across the entire year. They indicate that most customers buy more than one item.

Analyzing customer across all four charts of Figure 6, the business analyst may observe that the top dollar amount customers come back more frequently and buy more items. Month 12 has the largest number of customers. But months 3, 4, 5, and 9 have customers with the most sales. They frequently come back and purchase more items.
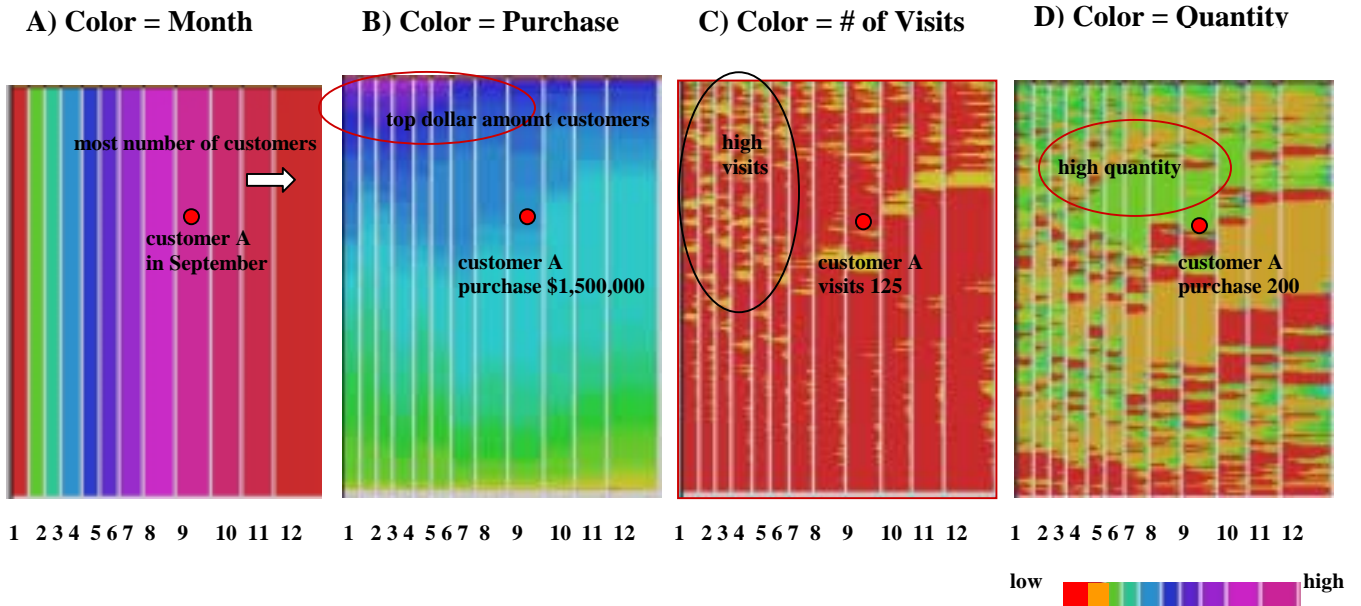
5

Figure 6: Pixel Bar Charts for mining over 150,000 E-Customer Purchasing Activities (by Month)

Each customer is represented by pixels that reside at the same location across multiple bar charts. From one location on each pixel bar chart we are able to find all the attributes of the pixel. The analyst can click on one pixel to get all the related information.

From Figure 6, we can observe the following information:

1. In September, customer A spent $1,500,000, visited 125 times, and bought 200 items.
2. The customers with most sales usually visited more often and purchased more products.
3. Most customers were one-time visitors.

6. Discussion

The integration of data mining with the visualization of e-customer behavior analysis is an emerging technology. At Hewlett-Packard laboratories, we have integrated a recent pixel bar chart technique into a visual data mining system. With this system, the user is able to switch between different algorithms, re-arrange the data, and re-visualize the results rapidly. We have used this system to visually mine datasets containing 3,500,000 customer web logs to analyze customer behavior of Internet usage.

From the results of these experiments, we have found that the pixel bar chart technique not only retains the simplicity of regular bar chart, but also allows the visualization of multiple attributes of a dataset. With linked multi-bar charts, the user can easily correlate different attributes to discover customer searching and purchasing patterns and trends. Further research is continuing on visual clustering and scalability issues.

Acknowledgements:

References:

[1] Daniel A. Keim, H. P. Kriegel, "VisDB: Database Exploration Using Muliti dimensional Visualization. IEEE Computer Graphics and Applications, Sep. 1994.
[2] Chris Stolte, Pat Hanrahan, "Ploaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases", InfoVis-2000, Salt Lake City, UT,, 2000.
[3] Stephen G. Eick, "Visualizing Multi-dimensional Data with ADVISOR/2000", VisualInsights, 1999.

[4] Daniel Keim, Ming Hao, Julian Ladisch, Meichun Hsu, Umeshwar Dayal, "Pixel Bar Charts: A New Technique for Visualizing Large Multi-Attribute Data Sets without Aggregation", Hewlett-Packard Laboratories Technical Report, HPL-2001-92, March 2001.

[5] Ming Hao, Umesh Dayal, Meichun Hsu, Bob D'eletto, Jim Becker, "A Java-based Visual Mining Infrastructure and Applications", InfoVis-99, CA USA, 1999.

[6] Alexander Hinneburg, Daniel A. Keim, Markus Wawryniuk: "HD-Eye: Visual Mining of High-dimensional Data", Computer Graphics and Applications, 1999. [7] Michael Ankerst, Christian Elsen, Martin Ester, Hans-Peter Kriegel, "Visual Classification: An Interactive Approach to Decision Tree Construction", KDD-99 San Diego, CA USA, 1999.

[8] Michael Ankest, Martin Ester, Hans-Peter Kriegel, "Towards an Effective Cooperation of the User and Computer for Classification", KDD 2000 Boston, MA USA, 2000.

[9] Daniel A. Keim, "Information Visualization Techniques for Exploring Large Databases", Tutorial Notes, Visualization 2000 Conf., Salt Lake City, UT, 2000.

[10] Daniel A. Keim: *'Designing Pixel-oriented Visualization Techniques: Theory and Applications',* TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS (TVCG) 2000.