Visual Mining Business Service Using Pixel Bar Charts

Ming C. Hao, Umeshwar Dayal, Fabio Casati
Intelligent Enterprise Technologies Laboratory
HP Laboratories Palo Alto
HPL-2004-112
June 29, 2004*

E-mail: {ming_hao, dayal_umeshwar, Casati}@hp.com

visualization, large volumes of data, pixel, color mappings, business operations

Basic bar charts have been commonly available, but they only show highly aggregated data. Finding the valuable information hidden in the data is essential to the success of business. We describe a new visualization technique called pixel bar charts, which are derived from regular bar charts. The basic idea of a pixel bar chart is to present all data values directly instead of aggregating them into a few data values. Pixel bar charts provide data distribution and exceptions besides aggregated data. The approach is to represent each data item (e.g. a business transaction) by a single pixel in the bar chart. The attribute of each data item is encoded into the pixel color and can be accessed and drilled down to the detail information as needed. Different color mappings are used to represent multiple attributes. This technique has been prototyped in three business service applications -- Business Operation Analysis, Sales Analysis, and Service Level Agreement Analysis at Hewlett Packard Laboratories. Our applications show the wide applicability and usefulness of this new idea.

* Internal Accession Date Only Approved for External Publication

Visual Mining Business Service Using Pixel Bar Charts

Ming C. Hao, Umeshwar Dayal, Fabio Casati
Hewlett Packard Research Laboratories, Palo Alto, CA
(ming_hao,dayal_umeshwar, Casati)@hp.com

ABSTRACT

Basic bar charts have been commonly available, but they only show highly aggregated data. Finding the valuable information hidden in the data is essential to the success of business. We describe a new visualization technique called pixel bar charts, which are derived from regular bar charts. The basic idea of a pixel bar chart is to present all data values directly instead of aggregating them into a few data values. Pixel bar charts provide data distribution and exceptions besides aggregated data. The approach is to represent each data item (e.g. a business transaction) by a single pixel in the bar chart. The attribute of each data item is encoded into the pixel color and can be accessed and drilled down to the detail information as needed. Different color mappings are used to represent multiple attributes. This technique has been prototyped in three business service applications – Business Operation Analysis, Sales Analysis, and Service Level Agreement Analysis at Hewlett Packard Laboratories. Our applications show the wide applicability and usefulness of this new idea.

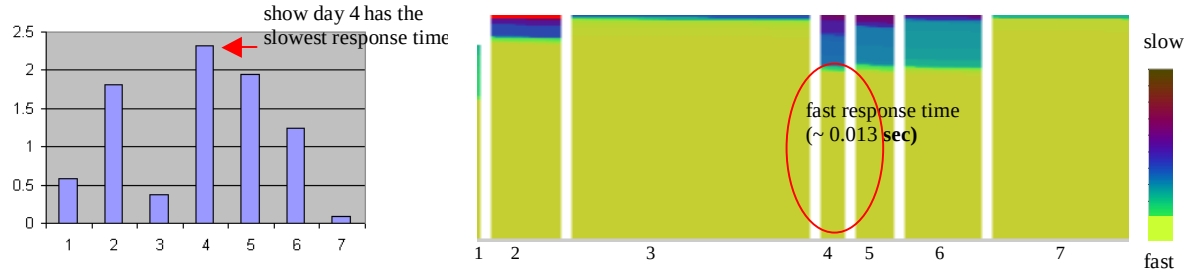
Keywords: Visualization, multi-attribute, Pixel Bar Charts, Business Operation Analysis, Web Service, Color Mappings

1. Motivation

The rapid growth of business information on the Internet has led to the availability of large volumes of data. Business research efforts have been focused on how to turn raw data into actionable knowledge. In order to find and retain customers, business analysts need to improve their business service quality based on prior information. In particular, it is crucial for companies to analyze their business service contracts and to reduce the risk of violation.

Basic bar charts have been commonly available, but only show highly aggregated data. Finding the valuable information hidden in the data is essential to the success of business. For example, in a weekly business performance report illustrated in Figure 1A, the analyst can easily find day 4 has the highest response time than other days. However, for analyzing large volumes of business transactions greater than 10,000, the visualization of aggregated data (i.e. average response time) is not sufficient. They are of limited value and are not able to show important information such as:

- data distributions of multiple attributes
- local patterns, correlations, and trends
- detailed information of an individual data item



(A) Bar Chart
(x-axis = day, y-axis = response time)

(B) Pixel Bar Chart
(x-axis = day, y-axis = response time, color = response time)

Figure 1: Business Service Response Time Analysis

To visualize web business services, other techniques include the tree map [1], OLAP data cubes [2] and business multi-dimensional Data [3]. They are emphasized in the areas of observing multi-dimensional business process flows and changes. Pixel bar chart differs from them by focusing on data distribution, patterns, and trends as described in the following sections.

2. Our Approach

In this paper, we describe a new visualization technique - pixel bar charts [4]. Pixel bar charts are derived from regular bar charts. The basic idea of a pixel bar chart is to present all data values directly instead of aggregating them into a few data values. The approach is to represent each data item (e.g. a business transaction) by a single pixel in the bar chart. The detailed information of one attribute of each data item is encoded into the pixel color and can be accessed and drilldown to the detail information as needed.

2.1. Construction of Pixel Bar Charts

Pixel bar chart integrates the idea of bar charts with an X-Y diagram. As illustrated in Figure 2, to construct a pixel bar charts, we specify the following attributes:

- dividing attribute (for between-bar partitioning, $\mathbf{D_x}$)
- ordering attribute (for outside-bar ordering, $\mathbf{D_y}$)
- ordering attributes (for within-bar ordering, $\mathbf{O_x}$, $\mathbf{O_y}$)
- coloring attributes (for pixel coloring, **C**).

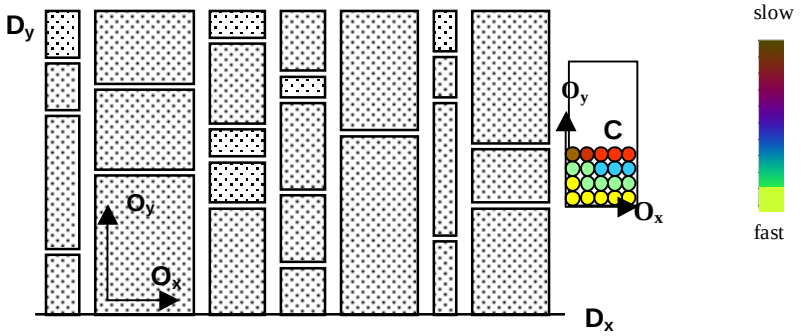


Figure 2: Definition of a Pixel Bar Chart
(e.g., D_x = day, D_y = process name, O_x = response time, O_y = quality value, C = I/O rate)

The pixel bar chart shown in figure 1B displays the same datasets as Figure1A. It uses *day* as the dividing attribute and *response time* as the x- and y-ordering attributes. The color represents the *response time* value of a business transaction. The fast response time corresponds to bright colors (yellow, green); the slow response time corresponds to dark colors (blue, burgundy, red). From the data distribution, we can discover that a majority number of fast business services occurred on day4 (large yellow area). In particular, it is crucial for analysts to realize that the number of transactions processed on day 4 is very small (narrow width bar). These valuable facts get lost from figure 1A. As a result, the business analyst may make a wrong decision.

2.2. Multiple Pixel Bar Charts & Correlation Mining

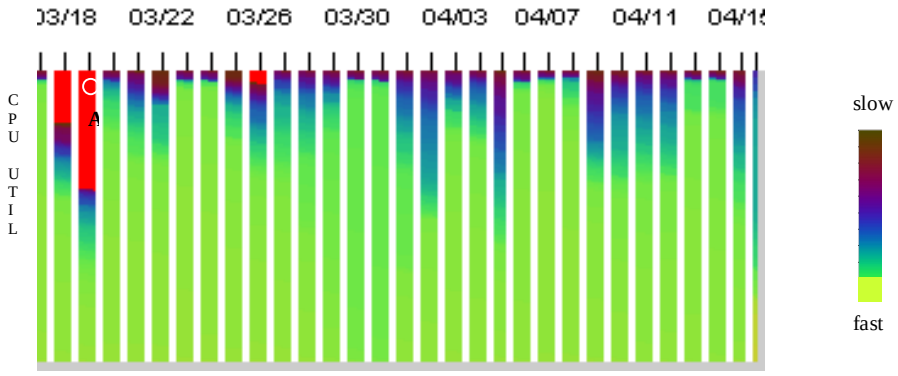
In many cases, the data to be analyzed consists of multiple attributes. Different color mappings are used to represent multiple attributes. The dividing and ordering attributes remain the same. In Figure 3A, the pixel bar chart shows business service CPU utilization activities in a month. Each pixel represents a business transaction record. The x-axis is the day slot (day); the y-axis is the %CPU utilization ordered from bottom to top. The colors in the different bar charts represent different CPU utilization values. The red represents transaction that exceeds a threshold of 99% CPU utilization. The analysts can obtain the following information on CPU utilization analysis:

a) CPU utilization exceed threshold (99%) is indicated by red color

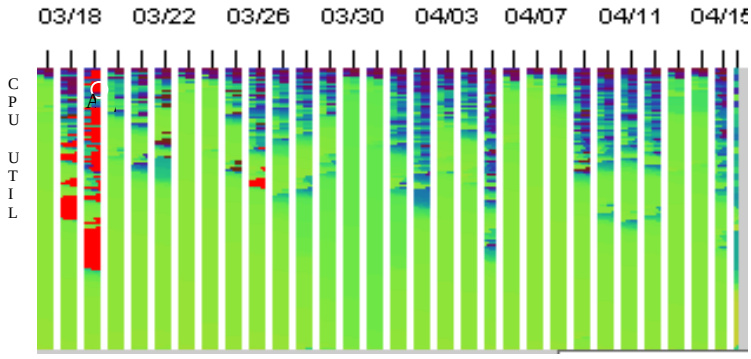
b) Business service has different CPU utilization during the month (Monday to Friday)
c) Business service has large CPU utilization during the weekdays and small CPU utilization during the weekends.
d) The highest CPU utilization occurred on 3/18 & 3/19 (Monday & Tuesday) with most red and blue colors.

Figure 3B illustrates the correlation between the CPU utilization and the I/O rate. The x-axis is the time slot (day); the y-axis is the I/O rate ordered from bottom to top. The colors represents different I/O rate. The red area represents the transactions which have high I/O rates exceeded the threshold. The analysts can make the conclusion that there is a close correlation between CPU utilization and I/O rate, because high I/O rates correspond to high CPU utilization as shown with darker color (red, blue, and burgundy) shown on the top band of the charts.

To identify correlations, a subset of data items in a pixel bar chart can be selected to get the pixels corresponding to related attribute values highlighted within the same display. A drill-down technique allows the viewing of all related information after selecting a single data item. When multi-bar charts are presented, pixels reside at the same location across all the charts with different attributes. In addition to discovering correlations and patterns, the user can select a single data item "A" (CPU utilization) in Figure 3 A to relate to other attribute value (I/O rate) in Figure 3B.



(A) CPU Utilization



(B) CPU utilization and I/O Rate Correlation

Figure 3: Business Service Multi-Attribute Correlations

3. The Pixel Bar Chart Visual Mining System

To analyze large volumes of transaction data with multiple attributes, pixel bar charts have been integrated with a data mining visualization system [5]. The system uses a web browser with a Java activator to allow real-time interactive visual data mining on the web.

3.1 Component Architecture

The pixel bar chart system connects to a data warehouse server and uses the database to query for detailed data as needed. The data to build the pixel array is kept in memory to support real-time manipulation and correlation. As illustrated in Figure 4, the pixel bar chart system architecture contains three basic components:

1. Placement
   One pixel represents one data record, i.e., a transaction. The pixel bar chart is partitioned based on attributes: e.g., day and process name. The ordering of pixels (y-axis) is based on attribute values, e.g., day in Figure 1B.

   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 response time in Figure 1B). 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 data warehouse as needed.

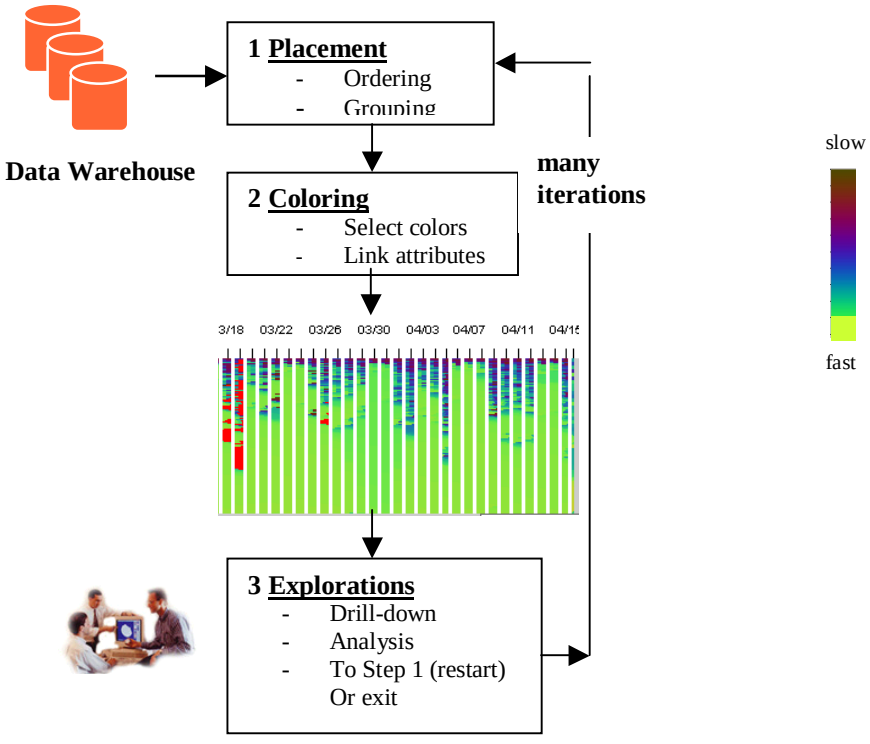


Figure 4: An Overview of Business Service Visual Mining System

3.2. User Interactions

Interactivity is an important aspect of the pixel bar chart system. To make large volumes of multi-attribute datasets easy to explore and interpret, the pixel bar chart system provides the following interaction capabilities: (1) visual querying; (2) layered drill-down; (3) multiple linked views; and (4) zoom.

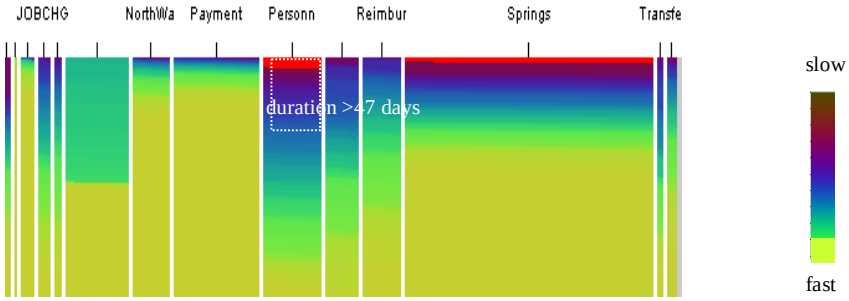
To mine large volumes of multi-attribute data, the user may want to try many different data arrangements. The attributes used for partitioning *(Dx, Dy)*, ordering *(Ox, Oy)*, and coloring *(C)* can be selected and changed at execution time. There are four types of pull down menus for the user to select input data and to construct pixel bar charts. The system provides a "Show" button for the user to, select, group, and visualize the pixel bar charts. The detailed information is displayed in the right lower corner of the window.

4. Applications

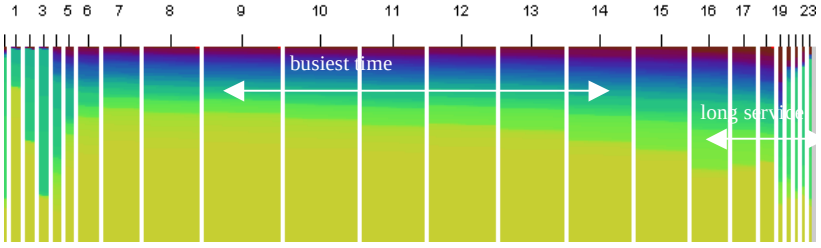
The pixel bar chart technique has been prototyped in three business process and business service applications – Business Process Distribution Analysis, Sales Analysis, and Service Level Agreement Analysis at Hewlett Packard Laboratories. Our applications show the wide applicability and usefulness of this new idea.

4.1. Business Process Distribution Analysis

Figure 5 illustrates 63,544 business process instances at HP web sites. Each pixel is a process instance.



(A) D_x = Process Name, O_y= duration, C = number of days)



(B) D_x = hour, O_y= duration, C = number of days)

Figure 5: Analysis of Business Process Duration by Day

Many important facts may be discovered in Figure 5 as flows:
1) In Figure 5A, there are multiple different business services (e.g. JOBCHG, Personnel, and Spring…). "Springs" has the most number of business instances with widest bar width. Most services duration are low (<1 day, color yellow) except "Personnel" (>47 days, colored dark blue and red)
2) The busiest time is during the hours between 9 and 14 as shown (with wider width bars) in Figure 5B. Late afternoon and evening have long running services (more blue and burgundy, less yellow from 17-23).
3) In both Figure 5A and 5B, large number of business service instances achieves good response time (color yellow, and green).

4.2. Sales Analysis

For sales analysis, sales specialists would like to discover new patterns and relationships in the invoice data. Common questions are "What is the sales growth rate in recent months?" , "Which product has the most sales? ", and "Where are the sales come from?"

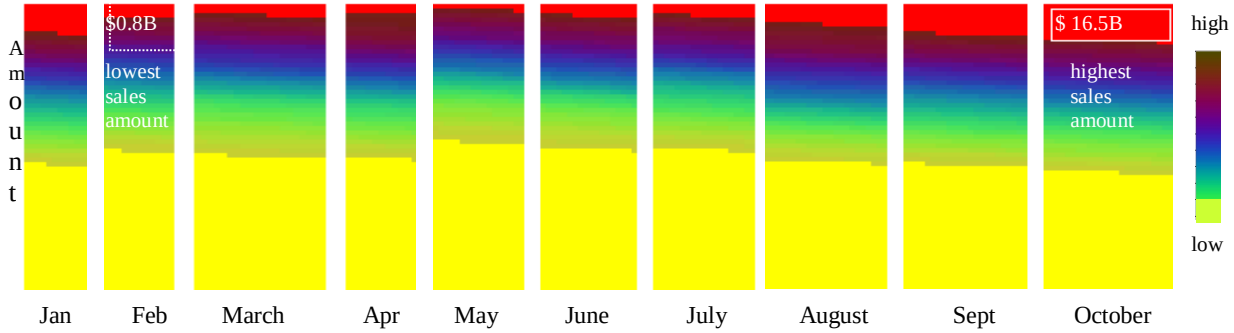


Figure 6: Sales Growth Rate Over 10 Months

In Figure 6, each pixel represents an invoice. Color represents sales amount. The red color indicates that the sales amount exceeds $1M. The analyst can find the following:

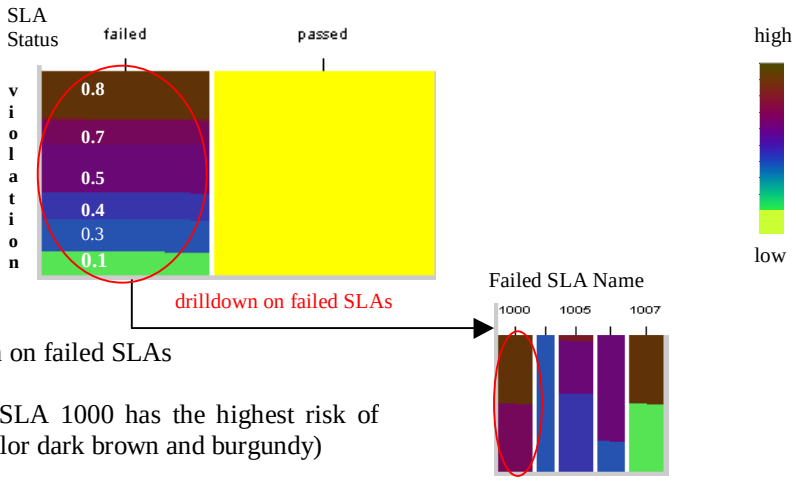
- Sales amount varies over time (shown by a growing color wave)
- Sales volume varies over time (shown by the width of a bar)
- Sales amount distribution over time (shown by the color – yellow, green, blue, burgundy, and red)
- Click on an invoice to find detail information on the sale (products, locations, time…)

The analyst is allowed to rubber-band an area inside a bar to display the total sale amount in the area. The sale amount represented by the rubber-band areas has grown from $0.8B (February) to $16.5B (October) as shown in Figure 6.

4.3 Service Level Agreement Analysis

One of the common questions e-business mangers ask is about service agreement performance. A critical area they would like to understand is which Service Level Agreements (SLAs) are being or close to being violated. Figure 7 shows a sample of pixel bar chart of 390,000 business service transactions with 986 SLAs. Each pixel represents an SLA transaction record. The ordering attributes are the SLA violation level from bottom to top and left to right. The color represents the value of violation level.

(A) Analysis of SLA Status (failed, passed)



(B) Drill down on failed SLAs

Shows that SLA 1000 has the highest risk of violation (color dark brown and burgundy)

Figure 7: Business Service Level Agreement Analysis

5. Conclusion

This paper presents a new pixel bar chart technique enabling analysts to visualize large multi-attribute business services. The approach is a generalization of traditional bar charts and x-y diagrams, that avoids the problem of losing information by aggregation. 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. Further research is continuing on dynamic color scalar and automatic zooming.

Acknowledgements

Many thanks to Professor Daniel Keim from University of Constance, Germany for his suggestions and ideas on pixel bar charts, to Kris Halvorsen, and Aad van Moorsel of HP Laboratories for their encouragement and suggestions, to Jerald T. Wade from Open View for providing suggestions, measurements, and data on business service, to Akhil Sahai and Vijay Machiraju for providing Business Contract Monitoring techniques and SLA data, to Ming-Chien Shan for information on the enterprise business management.

References

[1] Shneiderman B.: *'Tree Visualization with Treemaps: A 2D Space-Filling Approach',* ACM Transactions on Graphics, Vol. 11, No. 1, 1992, pp. 92-99.

[2] Stolte C., Tang D., Hanrahan P.: 'Multiscale Visualiztion Using Data Cubes', IEEE InfoVis99, Boston, Mass, 2002.

[3] Eick S.: Visualizing Multi-dimensional Data with ADVISOR/2000" VisualInsights, 1999.

[4] Keim D., Hao M., Dayal D., Hsu M., Ladisch J 'Pixel Bar Ch Charts: A New Technique for Visualizing Multi-Attribute Data Sets without Aggregation', IEEE/InfoVis2001

[5] Hao M, Dayal.U, Hsu M., D'eletto B., Becker J.'A Java-based Visual Mining Infrastructure and Applications', IEEE InfoVis99, San Francisco, CA. 1999.