Histographs: Interactive Clustering of Stacked Graphs

Pin Ren, Benjamin Watson
Northwestern University*

ABSTRACT

Visualization systems must intuitively display and allow interaction with large multivariate data on low-dimensional displays. One problem often encountered in the process is occlusion: the ambiguity that occurs when records from different data sets are mapped to the same display location. For example, because of occlusion summarizing 1000 graphs by simply stacking them one over another is pointless. We solve this problem by adapting the solution to a similar problem in the Information Murals system [2]: mapping the number of data elements at a location to display luminance. Inspired by histograms, which map data frequency to space, we call our solution *histographs*. By treating a histograph as a digital image, we can blur and highlight edges to emphasize data features. We also support interactive clustering of the data with data zooming and shape-based selection. We are currently investigating alternative occlusion blending schemes.

CR Categories: H.5.2 [Information Interfaces and Presentation: User Interfaces; I.3.3 [Computer Graphics]: Picture/Image Generation.

Keywords: information visualization, visualizing large data sets, data occlusion, digital image processing

1 INTRODUCTION

Visualizing large and multivariate data sets is an ongoing research effort within the information visualization community. Different techniques [1, 4] have been developed to address this type of visualization. The central problem is finding the overall trends and hidden patterns within these very large data sets, both by visual inspection and interactive grouping and clustering. When this data is graphed, many data points are mapped to the same location, hiding most of the data and greatly reducing the effectiveness of the visualization. The traditional solution to this occlusion problem is semi-transparency [5], which makes overlapping data objects translucent, blending them in depth order. This is effective if the number of occluding objects at any pixel is limited. Trutschl et al. [3] reposition occluded points using a smart jittering algorithm. This repositioning distorts the data and loses effectiveness when the number of data points exceeds the number of display pixels. Jearding and Stasko's overlooked solution in the Information Mural system [2] permitted large visualizations to be scaled down to small display sizes by mapping the number of data points at destination pixels to luminance.

2 HISTOGRAPHS

Inspired by the classic histograms that map data frequency to space, we propose *histographs*, which use the Information Mural frequency to luminance mapping to merge and summarize graphs. The result is a two-dimensional data texture that summarizes a stack of hundreds or thousands of graphs (see Figure 1). Building on this summary, our system permits users to group and cluster the graphs interactively, in the process gaining a better understanding of the graphed phenomena.

*{p-ren,watsonb}@cs.northwestern.edu

3 IMAGE-BASED TECHNIQUES AND INTERACTIONS

A histograph is in fact a digital image, enabling us to apply image processing techniques to improve the visualization. For example, a low pass, blurring filter emphasizes overall trends, while edge detectors highlight global data features (Figure 4). Moreover, the histograph itself can be thought of as a composite of hundreds or thousands of images of graphs. Many alternative compositing methods are possible, including maxima, minima, and schemes based on ungraphed data dimensions such as error, uncertainty and derivative.

User interaction is a key component of our system. In addition to filtering and compositing the graph stack, the user can zoom in on interesting graph sub-regions (Figure 3), select data with similar graphed shape using a lasso interaction (Figure 5), and reveal graph derivatives by mapping them to pixel hue (Figure 2). Users can also focus analysis on selected graphs, extracting them from the stack and visualizing them with their own histograph. In the process, users can build their own clustering of the data, improving their understanding of the recorded phenomena.

4 RESULTS

In order to test and demonstrate our idea, we used histographs to visualize a financial dataset containing one month of trading data for 1800 stocks. Figure 1 shows the histograph visualizing all of this data, plotted on a time vs. normalized log price graph. Notice the high-priced outlier, the many low-priced trades on the third day (August 5), and the daily price and volume discontinuities. Figure 2 shows the derivative coding, revealing colored vertical stripes that indicate strong daily price trends. Figures 3 and 4 zoom into the lower price range. In Figure 5, we show the use of the lasso to select and analyze a stock group that is rising in price.

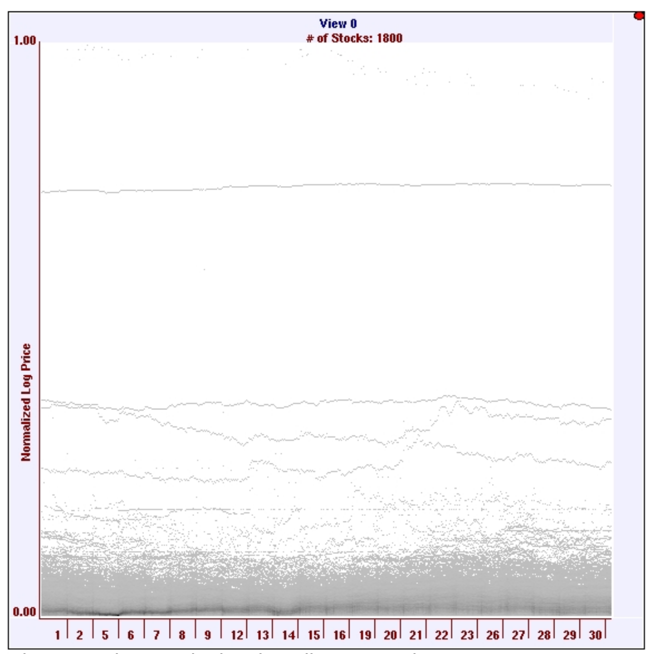

Figure 1: Histograph showing all 1800 stocks.

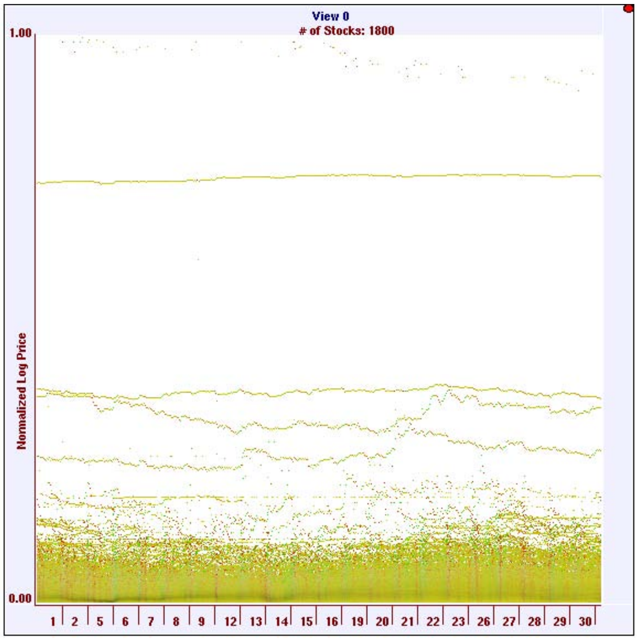

Figure 2: Histograph colored by first derivative. Green is positive slope, red negative; saturation is slope standard deviation.

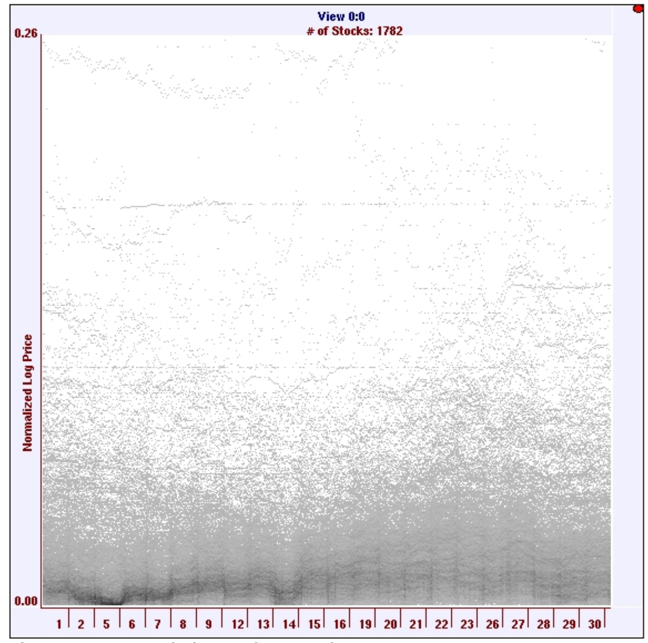

Figure 3: Zoomed view at lower price range.

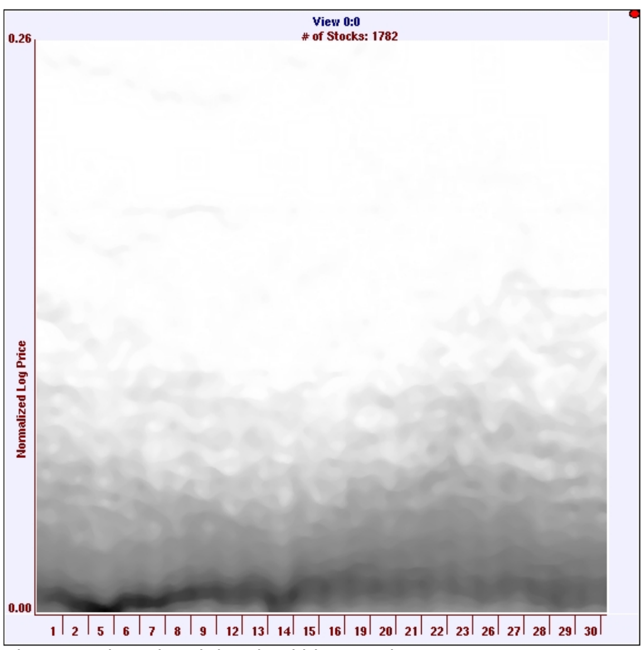

Figure 4: Blurred and denoised histograph.

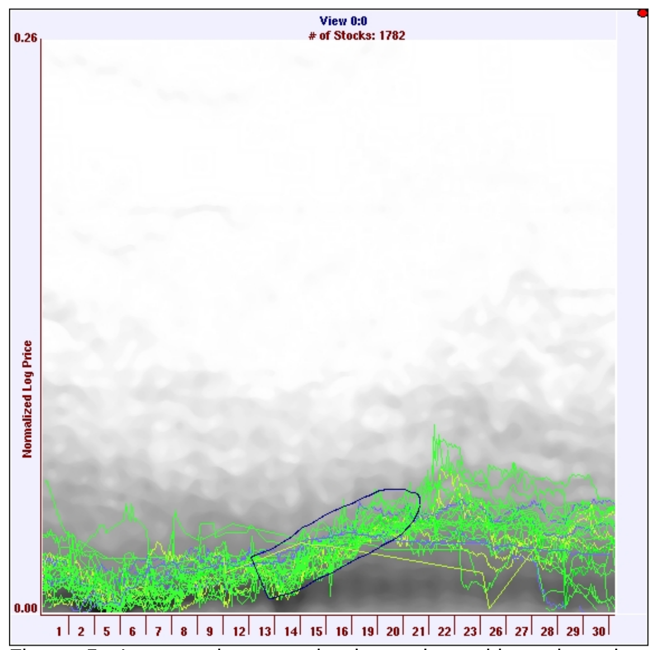

Figure 5: Lasso selects stocks increasing mid-month; colors indicate NYSE market category.

5 Future work

We are interested in other compositing methods, and have already experimented with using trade volume to weigh graphs. We are currently adding features to support interactive clustering, including methods for selecting and displaying groups of data. We are quite interesting in making histographs "live", using them to visualize data that is changing in real time. For our stock application, we are adding features for discerning overall patterns and validating models of those patterns. We will also examine how histographs might be applied to computer system performance traces and experimental sensor data.

References

[1] Jean-Daniel Fekete and Catherine Plaisant. Interactive information visualization of a million items. Proc. IEEE Info. Visualization, 117-124. 2002.

[2] Dean F. Jerding and John Stasko. The information mural: A technique for displaying and navigating large information spaces. *IEEE Trans. Visualization and Computer Graphics*, *4*, 3, ,257–271, 1998.

[3] Marjan Ttschl, Georges Grinstein, and Urska Cvek. Intelligently resolving point occlusion. Proc. IEEE Info. Visualization, 131-136.

[4] Martin Wattenberg. Visualizing the stock market. ACM CHI Extended Abstracts, 188– 189. 1999.

[5] Shumin Zhai, William Buxton, and Paul Milgram. The partial occlusion effect: utilizing semi-transparency in 3D human-computer interaction. *ACM Trans. Comp.-Hum. Interact.*, *3*, 3, 254–284, 1996.