

A Polyline-Based Visualization Technique for Tagged Time-Varying Data

Department of Computer Science Sayaka Yagi

1 Introduction

We have various interesting large scale time-varying data in our daily life (such as financial trading, and meteorological data), and they are commonly represented as polyline charts. Consequently, if hundreds or thousands of polylines are rendered in a single space, it may result in severe cluttering, and this may cause difficulty in reading data values and following data trends.

Several recent works have tried to address this problem. Hochheiser et al.[1] presented a gradient- and range-based query interface for polyline-based time-varying data visualization. Uchida et al.[2] presented a time-varying data visualization techniques, featuring sketch query on the clustered view. However, these techniques have dealt with polylines based on their data values or shapes, and do not use associated information.

Meanwhile, such information may be tightly correlated with the time-varying values. For example, air temperatures can be associated with weather attributes, blood pressure and electrocardiogram can be associated with descriptions of activities and so force. However, it is not always easy to determine the correlation between them. We believe it should be useful if time-varying data visualization techniques simultaneously display such associated information to assist the understanding of the correlation.

This thesis presents a new time-varying data visualization technique which supposes the associated information is attached as tags of the values of each time step.

2 Level-of-detail control for time varying data visualization

The technique supposes the following tagged time-varying data, consisting of a set of values $P = (p_1, p_2, \dots, p_n)$ represented as n polylines. we describe the tags of the i -th polyline as $w_i = (w_{i1}, w_{i2}, \dots, w_{im})$; w_{ij} denotes the tag at the j -th time of the i -th polyline, and the values of a polyline as $p_i = (p_{i1}, p_{i2}, \dots, p_{im})$; p_{ij} denotes the value at the j -th time of the i -th polyline. We draw the set of values as a polyline chart, while the horizontal axis denotes the 1st to the m -th time, and the vertical axis denote the magnitude of the values.

2.1 Clustering and representative polyline selection

The procedure of quantization and clipping step is as follows.

1. Generate a grid covering the drawing area, and then divides into $a \times b$ subspaces(as shown in Figure 1(a)).

2. Sample P at the time t_0 to t_a , and temporarily quantize the sampled values at b_0 to b_b . Here, b_i is the value at the i -th horizontal line of the grid.
3. Generate groups of polylines, if the polylines have the same quantized values both at t_{i-1} and t_i (as shown in Figure 1(b)).
4. Clip polylines of a group by t_{i-1} and t_i and generate clusters of the clipped polygons (as shown in Figure 1(c)).

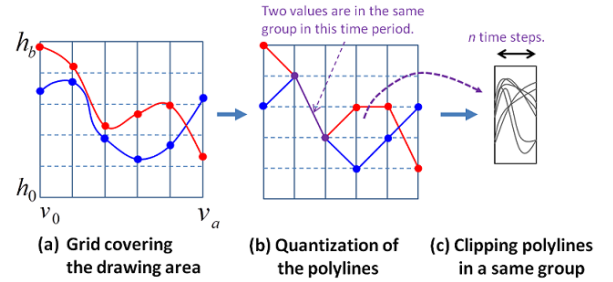


Figure 1: Quantization and clipping of polyline.

Figure 2 denotes the procedure, while the colors of polylines denote their dominant tags. Here, the technique regards the clipped polylines as n -dimensional vectors, while they contain n time steps between t_{i-1} and t_i . Procedure for clustering of tagged polylines is as follows.

1. Divide the clipped polylines according to tags, using a dendrogram from the polylines constructed according to similarities of their n -dimensional vectors $(w_{jt_{i-1}}, \dots, w_{jt_i})$ (as shown in Figure 2(b)).
2. Apply a non-hierarchical clustering (e.g. k-means) to the polylines in each cluster, using their n -dimensional vectors $(p_{jt_{i-1}}, \dots, p_{jt_i})$, and generate clusters consisting of similarly tagged and shaped fragments of polylines (as shown in Figure 2(c)).
3. Select representative polylines for each cluster (as shown in Figure 2(d)).

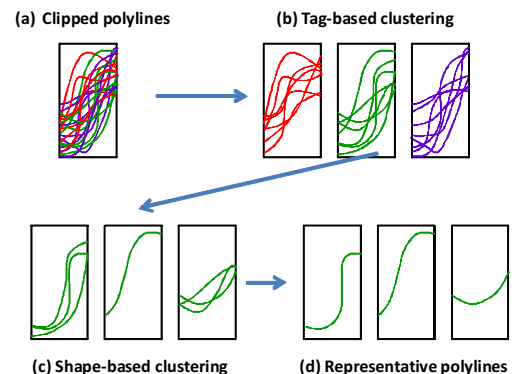


Figure 2: Clustering of polylines.

2.2 Interactive Visualization

The technique represents the time-varying data as colored polylines. It assigns colors to the tags and draws the polyline in the assigned colors. If the vertices of a segment of a polyline have different tags, it interpolates the colors along the segments.

Initially our technique draws only the representative polylines. Our current implementation generates several clustering results, with several configurations of the grid and the clustering process. Smoothly replacing the clustering results, our technique seamlessly displays several levels of numbers of representatives.

The technique also provides click and sketch interfaces, so that users can specify interesting representatives by directly clicking or sketching particular shapes. When a user clicks a point on the display, the technique calculates distances between the point and all segments of the drawn polylines. If at least one of the segments of a polyline is enough close to the clicked point, the technique highlights the current polyline. When a user draws a curve on the display, the technique samples several points on the curve, and calculates distances between the sampled points and all segments of the drawn polylines. If at least one of the segments of a polyline is enough close to each of the sampled points, the technique highlights the current polyline as shown in Figure 3.

Users can specify particular tags to be extracted by the above query operations. It can highlight only the parts of the polylines corresponding to the specified tags while the click or sketch operations. Also, it can reactivate only the parts of the non-representative polylines corresponding to the specified tags.

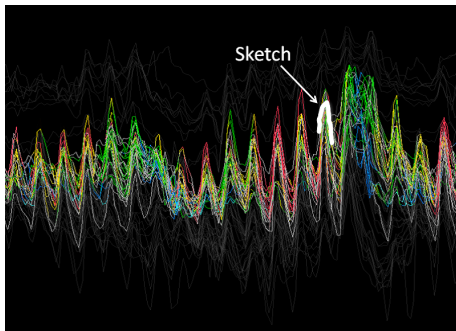


Figure 3: Sketch interface.

3 Examples

We applied Japanese weather data recorded by AMeDAS (Automated Meteorological Data Acquisition System) to the presented technique. We extracted time-varying temperature data observed at 376 points in every 3 hours. We then assigned weather tags including "Clear", "Sunny", "Cloudy", "Rainy", and "Snowy" to temperature value of each time of each point.

Our implementation draws the segments of "Clear" polylines in red, "Sunny" in yellow, "Cloudy" in

green, "Rainy" in blue, and "Snowy" in cyan. Exceptionally it draws the segments in gray if we cannot obtain the weather data. While using click or sketch interfaces, our implementation draws selected polylines brightly, and others in gray.

Figure 4 shows a zoom up view of temperature variation during five days. Figure 4(a) shows the original view of the tagged temperature data before applying level-of-detail control, and Figure 4(b) shows after applying level-of-detail control with tag-based clustering. Figure 4(a) has many overlaps of polylines, especially non-tagged ones colored in gray. While Figure 4(b) represents the features of the data more clearly. In Figure 4(b), we can find that extremely lower temperature can be often observed from "Clear" or "Sunny" points. It is difficult to obtain such knowledge from the visualization results with shape-based clustering without considering tags, as shown in Figure 4(c).

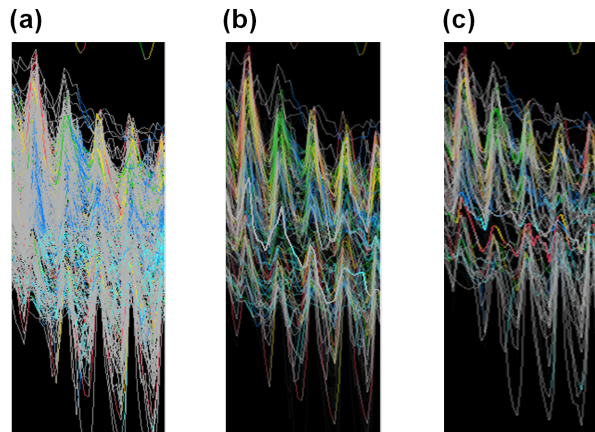


Figure 4: Zoom up view: (a) without level-of-detail control. (b) with level-of-detail control (tag-based clustering). (c) with level-of-detail control (shape-based clustering).

4 Conclusion

This thesis presented a polyline-based visualization technique for tagged time-varying data. By using the technique, we can easily observe overall features without missing detailed features of tags, while reducing cluttering among polylines.

The following issues will be our future work:

- Discuss more effective representation for:
 - more kinds of (e.g. more than 10) tags.
 - multiple tags at a point.
- Add features to the technique so that we can focus on time-varying value variations with particular patterns of tag changes.

Reference

- [1] H. Hochheiser, B. Shneiderman, Dynamic query tools for time series data sets: Timebox widgets for interactive exploration, *Information Visualization*, 3(1):1-18, 2004.
- [2] Y. Uchida, T. Itoh, A visualization and level-of-detail control technique for large scale time series data, *13th International Conference on Information Visualization*, 80-85, 2009.