STFMap: Query- and Feature-Driven Visualization of Large Time Series Data Sets*

K. Selçuk Candan
Arizona State University
Tempe, AZ 85287, USA
candan@asu.edu

Rosaria Rossini
University of Torino
Torino, Italy
rossini@di.unito.it

Maria Luisa Sapino
University of Torino
Torino, Italy
mlsapino@di.unito.it

Xiaolan Wang
Arizona State University
Tempe, AZ 85287, USA
xwang220@asu.edu

ABSTRACT

Since many applications rely on time-based data, visualizing temporal data and helping experts explore large time series data sets are critical in many application domains. In this interactive system preview, we argue that time series often carry structural features that can, if efficiently identified and effectively visualized, help reduce visual overload and help the user quickly focus on the relevant portions of the data sets. Relying on this observation, we introduce a novel *STFMap* system, which includes four innovative query- and feature-driven time series data set visualization techniques: (a) *segment-maps*, (b) *warp-maps*, (c) *stretch-maps*, and (d) *feature-maps*. These rely on the *salient temporal features* of the time series and their alignments with respect to the given user query to help users explore the data set in a query-driven fashion.

Categories and Subject Descriptors

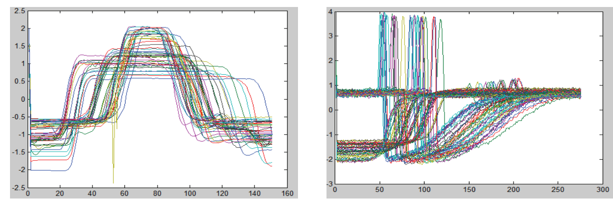
H.5 [**Information Interfaces and Presentation**]: Miscellaneous

Keywords

Time series data sets, data exploration

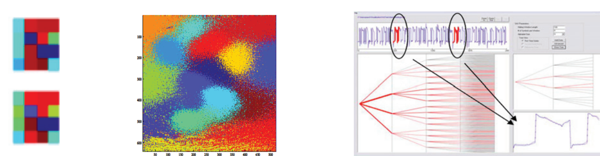
1. INTRODUCTION

Since many applications rely on time-based data, visualizing large temporal data sets is critical in many application domains. As can be seen in Figure 1, naive plot-based visualization of large time series data sets can become overly dense and, as a result, subtle differences and similarities between the time series can be lost in the resulting visual clutter. While visualization aids, such as time boxes [3], have been proposed to help reduce visual clutter in such plots, it is still hard to effectively compare features of the time series at different scales as the visualization is dominated by features that are large in amplitude. Various alternative visualization techniques, such as SAX-based time series bitmaps [6] or dimensionality-reduction based techniques [1], try to address this challenge by mapping time series to representations that are good at capturing the global characteristics of the data at the cost of loosing local features (Figure 2(a,b)). Other techniques, such as



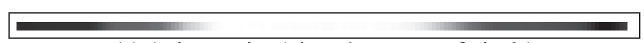
(a) Gun data set (b) Trace data set

Figure 1: Plots of sample the data sets in [5]

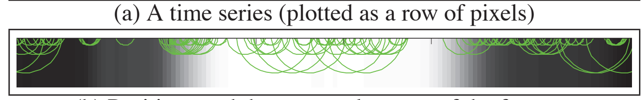


(a) Two sample time series bitmaps (b) Data set visualization by dimensionality reduction (c) VizTree (of a single time series)

Figure 2: (a) Two time series bitmap examples taken from [6]; (b) dimensionality reduction based visualization of an entire time series taken from [1] (each time series is represented by a single pixel); (c) a sample VizTree (taken from [7])



(a) A time series (plotted as a row of pixels)



(b) Positions and the temporal scopes of the features

Figure 3: (a) A time series visualized as a row of pixels and (b) the features identified on the series – each half circle is a feature

VizTree [7], try to summarize both the global and local structures of a given time series by encoding the series in representations, such as trees, in which the frequency and other properties of patterns can be mapped onto visual properties (Figure 2(c)). Yet, these also face difficulties in representing both high-level and low-level structures in the data set.

In [2], we recognized that time series data often carry structural evidences (in the form of *salient temporal features*) that can be used for indexing and retrieval and we proposed a *SIFT-like*[1] `sDTW` algorithm to identify robust points on a given time series (Figure 3). While the details are outside of the scope of this paper, it is impor-

*This work is supported by the NSF Grant 1043583 " MiNC: NSDL Middleware for Network- and Context-aware Recommendations". Authors are listed in alphabetical order.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
ACM 978-1-4503-1156-4/12/10.

[1]The original scale-invariant feature transform (SIFT) [8] algorithm is for 2D images.

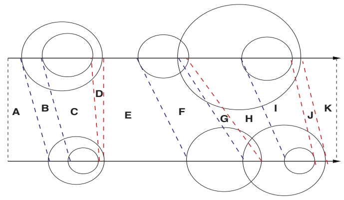


Figure 4: Segments defined by the temporally consistent feature matches across two time series

tant to state that the identified salient features (defined by a center and a temporal radius/scope) are robust against noise and common transformations, such as temporal shifts. Moreover, scale invariance enables the extracted salient features to be robust against variations in speed of time.

In this paper we further recognize that if efficiently identified and effectively leveraged within the context defined by a user query, these salient features can also help user quickly focus on the relevant portions of the data sets. Relying on this observation, we describe an interactive system demonstration of a *STFMap* system, which includes (among others) four novel temporal data set visualization techniques, (a) *segment-maps*, (b) *warp-maps*, (c) *stretch-maps*, and (d) *feature-maps*. All four techniques leverage the *salient features* of the time series, identified through a preprocessing step, to help explore large temporal data sets within the context of a given query. In the rest of the proposal, we describe the interactive system preview of the *STFMap* system.

2. INTERACTIVE SYSTEM PREVIEW

The user first selects, using the GUI interface, a data set, D, and provides a query series, q, with the goal of exploring the time series in D based on their similarities and differences to q. We propose to demonstrate *STFMap* using data sets from [5] (Figure 1) as well as economic index time series obtained from [9].

2.1 Demo I: Segment-Map Visualization

STFMap first analyzes the feature alignments between the query and all time series in D and identifies temporally consistent matching pairs of salient features across q and series in D. As illustrated in Figure 4, given two series q and $s \in D$, the boundaries of the matching features partition both q and s into segments: consecutive segments differ in that some existing features have ended or new features have started. Intuitively, the segment boundaries of s describe how the features of s relate to the features of q. Therefore, a segment-map describes how the segments of the time series in D, *relative to* the query time series q, are distributed.

Figure 5(a) shows a segment-map, where 50 time series in the `Gun` data set are plotted as *columns* of pixels (blue indicates low amplitude points, whereas red indicates high amplitude points). Segment boundaries are indicated with horizontal lines.

The *segment-map* also enables the user to have a multi-resolution view of the distribution of matching features: one can roll-up the data by clustering nearby features and the segments they induce. The boundaries of the rolled-up features define a new smaller set of segments, where some of the consecutive segments are combined into larger segments. Figure 5(b) shows a rolled-up segment map; the figure also shows a *drill-down window* on the segment-map.

2.2 Demo II: Warp-Map Visualization

Dynamic time warping (DTW [2, 4]) represents the minimum time warping needed to align to series. The demonstration will also include warp-maps, which visualize the warping needed for aligning the time series in D with a given query series, q.

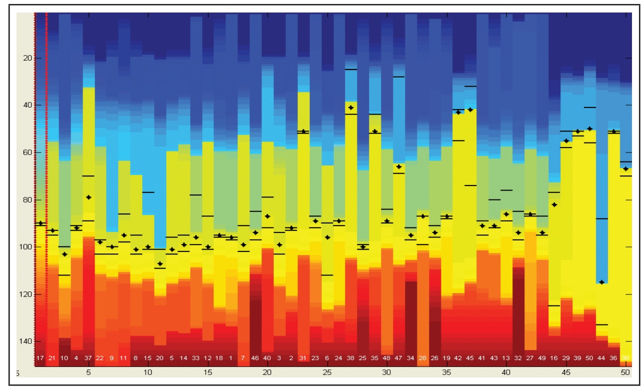


Figure 6: A warp-map (`Gun` data set): the query series, q, is plotted with colors changing from blue ($t = 0$), green ($t \sim 75$), to red ($t = 150$); each time point on each data series in the data set is painted with the color of the corresponding time point on the query series. In this example, on each time series, the system also highlights the point (and the enclosing segment boundaries) corresponding to a point-of-interest on q. Time series are ordered left-to-right according to their DTW distances from the query series (leftmost column).

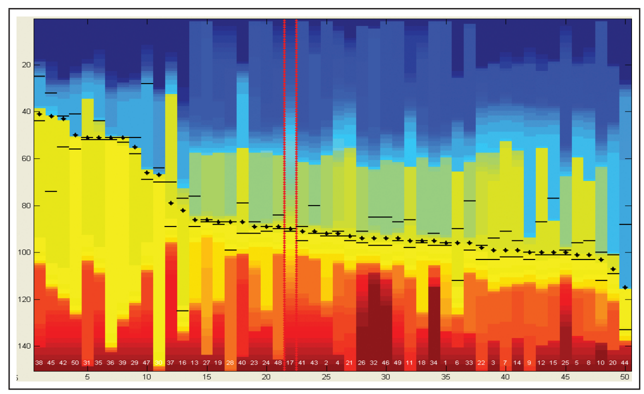
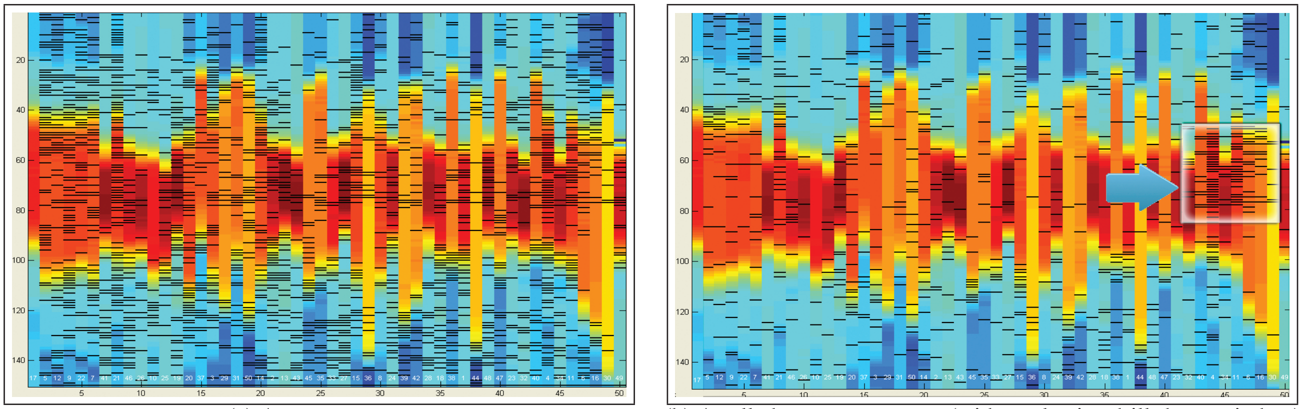


Figure 7: A warp-map (`Gun` data set): Unlike in Figure 6, the series are ordered based on the matching positions of the interest-point – the query series, q, is highlighted in a red box

Warp-maps are created as follows. The query series, q, is plotted with colors varying from blue (for $t = 0$), green (for $t \sim 75$), to red (for $t = 150$). Let $s \in D$ be a data series and q be the user's query; the warp path maps each time point $s[i]$ of s to one or more consecutive time points, denoted $q(s[i])$, on q. The point $s[i]$ is, then, painted with the color of corresponding point, $cor(q, s[i])$, which is the average of the time points in the set, $q(s[i])$.

As illustrated in Figure 6, the resulting warp-map helps the user observe the warping needed to align the series in the data set with the query series q. Note that, unlike the data map, where the amplitude of the data plays a significant role, in the warp-map the user can directly observe the temporal alignments among the series.

The warp-map in Figure 6 also includes a point of interest selected by the user on the query series and the corresponding points (and the enclosing segments) highlighted on each data series. This point-of-interest driven exploration helps the user to observe the amount of stretching or warping local to the point of interest. As shown in Figure 7, the point-of-interest based visualization can be further strengthened by ordering the time series in D based on the positions of the matching interest points instead of based on the DTW-distance to the query.



(a) A segment-map

(b) A rolled-up segment-map (with a selective drill-down window)

Figure 5: (a) The segment-map of the `Gun` data set (blue indicates low amplitude points on the series, red indicates high amplitude points– horizontal bars correspond to segment boundaries); (b) an rolled-up with a selective drill-down. Time series are ordered left-to-right according to their DTW distances from the query series (leftmost column)

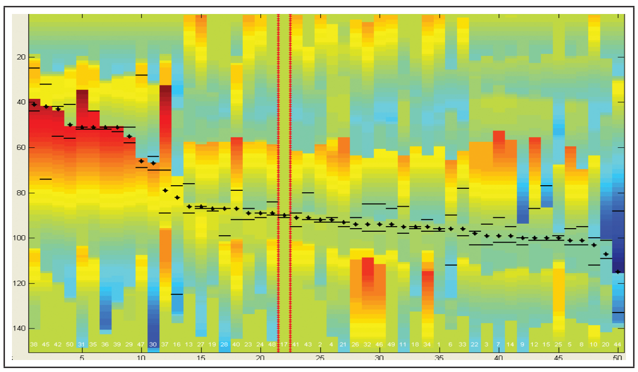


Figure 8: A stretch-map (`Gun` data set): red indicates large forward stretching, dark blue indicates large backward stretching, and green corresponds to low or no stretching)

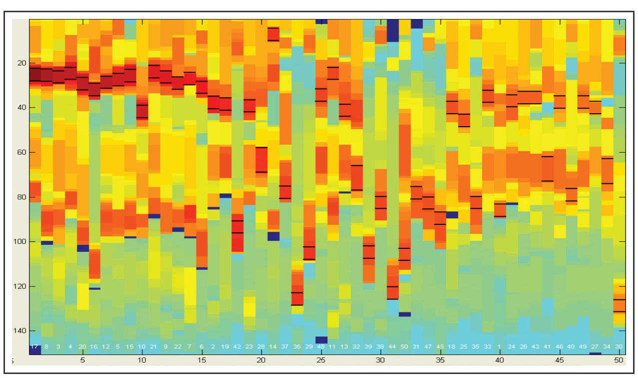


Figure 9: A feature-map (`Gun` data set): darker red indicates better feature matches; series are ordered left-to-right in the order of best matching feature and on each series the boundaries of the best matching feature is highlighted

2.3 Demo III: Stretch-Map Visualization

The demonstration will also include novel stretch-maps which, while being similar to warp-map, help the user focus on the major alignment differences more quickly: see the red and blue regions at the left and right extremes of the stretch-map shown in Figure 8. In a stretch-map, the point $s[i]$ is painted with the value of $i - cor(q, s[i])$: in this case, the more negative the difference between the matching positions is, the darker the blue and the more positive the difference is, the darker is the red used for painting $s[i]$. Shades of green denote none or little stretching.

2.4 Demo IV: Feature-Map Visualization

The *STFMap* system also includes visualization techniques to help users directly explore the data set based on a selected feature of interest on q. Let f be a salient temporal feature of interest on the query. For each point $s[i]$ on series s, *STFMap* first identifies the set, $\mathcal{F}(s[i])$, of features covering $s[i]$. It then computes a similarity score [2], $\mu(f, f')$ for each $f' \in \mathcal{F}(s[i])$ and paints the point $s[i]$ with $max\{\mu(f, f') \mid f' \in \mathcal{F}(s[i])\}$. Intuitively, each point is painted according to the matching degree of the best matching feature it represents. As shown in Figure 9, on the resulting feature-map, the user can quickly observe the positions of the likely matches to the feature of interest, f. We will also present the use of feature-maps during the demonstration of *STFMap*.

3. CONCLUSIONS

Recognizing that time series often contain *salient temporal features* that can, if efficiently identified and effectively leveraged, help users quickly focus on the relevant portions of a data set, we propose to demonstrate the *STFMap* system and its novel *salient feature* based temporal data set visualization techniques, (a) *segment-maps*, (b) *warp-maps*, (c) *stretch-maps*, and (d) *feature-maps*. These rely on the properties of the salient features extracted from the time series in the data set and their alignment with respect to the features of the query series of interest.

4. REFERENCES

[1] A. Bagnall *et al*. A bit level representation for time series data mining with shape based similarity. Data Min. Knowl. Discov. 13, 1, 11-40, 2006.

[2] K.S. Candan, R. Rossini, M.L. Sapino, X. Wang. sDTW: Computing DTW Distances using Locally Relevant Constraints based on Salient Feature Alignments. VLDB 2012.

[3] E.Keogh, H.Hochheiser, and B.Shneiderman. An augmented visual query mechanism for finding patterns in time series data. FQAS2002.

[4] E. Keogh. Exact indexing of dynamic time warping. VLDB. pp 406-417, 2002.

[5] E.Keogh *et al*. The UCR Time Series Classification /Clustering Homepage: www.cs.ucr.edu/∼eamonn/ time_series_data/ 2011.

[6] N. Kumar, V.N. Lolla, E.Keogh, S. Lonardi, C.A. Ratanamahatana, and L. Wei. Time-series bitmaps: a practical visualization tool for working with large time series databases. SDM 2005.

[7] J. Lin *et al*. Visualizing and discovering non-trivial patterns in large time series databases. Information Visualization 4, 2, 61-82, 2005.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[9] Paola Pisano, Marco Pironti. Personal communication. 2011.