Multidimensional Projections to Explore Time-Varying Multivariate Volume Data

Christian Wong, Maria Cristina. F. Oliveira, Rosane Minghim
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
São Carlos/SP, Brazil
{cwong,cristina,rminghim}@icmc.usp.br

***Abstract*—Multidimensional Projections (MPs) have become popular as visual data analysis tools in several application domains, including Scientific Visualization. Current techniques are fast, precise and capable of handling local and global data features, having successfully supported spatial and abstract data visualizations. However, two major shortcomings hinder their application for exploratory analysis of time-varying multivariate volumetric data. Current techniques lack visual coherence when applied to data collected across consecutive time stamps and offer little support to investigating attribute-specific questions. Both are relevant properties when analysing time varying volumes. In this paper we revisit projection methods from this perspective and introduce modifications into two existing high-performance techniques to ensure temporal coherence. We also propose a hybrid visualization strategy that can assist users investigating the role of a specific attribute on data behavior through time. We illustrate how our approaches enhance projection-based visual exploration of time-varying multivariate volume data with their application to data sets from three distinct simulations, made available for editions of the *IEEE Visualization Contest*.**

***Keywords*-Visualization; Scientific Visualization; Multidimensional Projections; Exploratory Volume Visualization.**

I. Introduction

Volume visualization [1], [2] deals with methods to explore, analyze and visualize volumetric data, supporting the investigation of physical phenomena. Volumetric data sets are acquired by sampling, simulation or modeling techniques and are typically represented as three-dimensional meshes. Such spatial representations register observations of multiple scalar and vector variables acquired or computed over time and associated to voxels or cell nodes.

Understanding and exploring time varying multivariate volume sequences is still a major challenge to scientists. In this paper we investigate the role of tools based on multidimensional projections (MPs) to assist their exploration. Our hypothesis is that MPs are interesting in this scenario due to their ability to convey global relationships among the multivariate cells, particularly in application domains described by multidimensional data. Two critical issues impair usage of MPs in this context:

1) Few available projection techniques can handle large volumes in a time-varying context.
2) Projections offer little support to users investigating the role of specific attributes in global behavior.

Addressing the above issues requires deriving MP techniques capable of handling time-varying multivariate volumes in real-time, a scenario that imposes the following requirements: (i) speed, i.e., multiple volumes must be projected at interactive rates; (ii) layout quality, i.e., the 2D layout must preserve the feature space description so that groups formed in original space are observable in the projected space; (iii) temporal coherence, i.e., a sequence of time-stamped projections of a volume should preserve global organization and orientation of elements, so that users can maintain a mental model while focusing on the relevant changes.

Fast techniques such as Fastmap [3] and PLP (the *Part-Linear Projection*) [4] could be considered, as they handle large high-dimensional data at interactive rates. However, they do not guarantee temporal coherence when projecting a sequence of time-stamped volumes. We thus modified their original formulation to produce time-coherent layouts. Based on their novel formulations we propose a hybrid 2D visualization that combines a 1D projection of a feature sub-space with a 1D scatterplot of a focus attribute, as an additional functionality to support investigating the role of specific attributes on global data behavior.

II. Related Work

MPs are feature-based point placements that can support exploratory tasks aimed at investigating global and specific data behavior. In a good layout, voxels with similar properties are mapped as elements placed in neighboring areas. Several authors describe exploratory volume visualization solutions that integrate projection layouts or other multidimensional visualizations with classical scalar or vector volume visualizations, typically combining feature space layouts with standard object space views of the volume data.

Chen et al. [5], for example, provide linked 3D and 2D views to assist exploration of *Diffusion Tensor Imaging* (DTI) fibers. A 2D layout is computed with *Classical Scaling* [6] that preserves the spatial relationships of the fiber tracts and removes the visual clutter characteristic of the 3D visualization. For a similar application Jianu et al. [7] employ a force direct method [8], adopting the acceleration technique by Chalmers [9] to achieve interactive performance. Poco et al. [10] use a fast supervised MP called LAMP [11] to generate layouts that support exploratory visualization of large fiber

1530-1834/13 $26.00 © 2013 IEEE
DOI 10.1109/SIBGRAPI.2013.24

tracking data sets, offering interaction functionalities to select or change groups of bundled fibers for further inspection.

Daniels et al. [12] employ MPs to assist interactive exploration of vector fields, observing that similar features within the vector field, even if placed spatially apart, may share similar neighborhoods in the feature space. The projection layouts favor the perception of these neighborhoods. Users can move control points and design a texture over an interactive canvas to improve layout quality.

Santos et al. [13] describe a visualization strategy to support identification of vortices in vector data. Authors compute streamlines from velocity data, extract shape information as features and project them using the PLP multidimensional projection [4].

Our contribution shares similarities with that of Guo et al. [14], who embed MP layouts in a Parallel Coordinates plot to define transfer functions for volume rendering. The approach has been designed to be scalable and deliver good performance on data sets of varying sizes and dimensionalities [15], but temporal coherence is not preserved when observing multiple time steps.

In fact, most existing MP techniques do not ensure temporal coherence. To the best of our knowledge, the only MP applicable to general high-dimensional data and capable of preserving temporal coherence is the *Part-Linear Multidimensional Projection* (PLMP) [16]. This is a fast and precise technique targeted at supporting interactive exploration of massive data sets and also handling time-varying volumes. We take it as the baseline to compare the performance and behavior of the techniques introduced in this paper. We show that they exibit comparable or superior performance relative to certain properties relevant to time varying volume exploration.

III. Background

This section introduces the original formulations of Fastmap and PLP, which will be later modified to ensure time coherence and better visual grouping of elements when projecting large time-varying data sets.

A. Fastmap

Fastmap [3] is a fast algorithm for mapping high-dimensional objects (data instances) into a lower k-dimensional space. It attempts to preserve distances (or dissimilarities) among data objects while projecting them onto k mutually orthogonal directions. The rationale is to recursively project the objects onto k hyperplanes defined by a subset of the dimensions to obtain the object coordinate in each projected dimension. If $k = 2, 3$ the mapping may be interpreted as a projection technique.

At each recursion step a projection direction is defined. Within each hyperplane, two pivot objects (O_a and O_b) are selected and then all objects are projected on the line defined by the pivots. The projected coordinates of an arbitrary object O_i in each hyperplane is computed using the Cosine Law:

$$d_{b,i}^2 = d_{a,i}^2 + d_{a,b}^2 - 2x_i d_{a,b} \tag{1}$$

where $d_{i,j}$ denotes the distance between any two objects O_i and O_j, x_i is the desired projected coordinate of O_i in the line $\overline{ab}$ that connects the pivots O_a and O_b. Equation 1 may be solved for x_i and the distance between any two projected objects may be computed employing the Euclidean distance:

$$d'_{i,j} = \sqrt{d_{i,j}^2 - (x_i - x_j)^2} \tag{2}$$

Equation 2 allows to project over another line which is orthogonal to the first.

A final issue is how to select the pair of pivots. The ideal choice of picking the two data objects farthest apart is computationally expensive. The following heuristic [3] provides an alternative. First an arbitrary object A is chosen. Then the object farthest apart from A is found and taken as the first pivot, O_a. Again, the object farthest apart from O_a is found and taken as the second pivot, O_b.

However, the selected pivots are not necessarily good representatives of the entire data set, as a data point P may fall outside the area they define. This affects the method's precision, which may be improved by running further iterations of the heuristic, taking the second pivot as the initial object. Naturally, this solution incurs in higher computational cost.

The computational cost of Fastmap is determined by the two steps executed as many times as the number of target dimensions. Choosing the pivots requires three iterations over the data, resulting in complexity $O(3n) = O(n)$, for n data points. The second step is just an iteration over the data to compute the projected coordinates using the cosine law. Thus, overall cost is $O(k(3n + n)) = O(n)$, since k is low (1, 2 or 3).

B. Piecewise Laplacian-based Projection

The Piecewise Laplacian-based Projection (PLP) [4] is fast enough to support real-time interaction with large data sets. The method comprises three major steps: sampling, building a neighborhood graph and solving a Laplacian linear system.

The sampling step selects a small representative subset from the original data, and may be performed automatically by taking samples from clusters obtained from the data. These samples are projected with a *Multidimensional Scaling* (MDS) technique such as the *Least Squares Projection* (LSP) [17]. Then, the neighborhood relations are defined by a *k-nearest neighbor graph (kNNG)*.

Finally, a Laplacian linear system is built based on the assumption that each data object O_i can be written as a convex combination of its nearest neighbors in the visual space. Under this hypothesis, the projected coordinates of the data objects may be written as:

$$x_i = \sum_{O_{i_j} \in kNN(O_i)} \alpha_{ij}(x_{O_{i_j}}, y_{O_{i_j}}) \tag{3}$$

where $\alpha_{ij} > 0$ and $\sum \alpha_{ij} = 1$.

From Equation 3 a linear system may be assembled as: $Lx = 0$ (and likewise $Ly = 0$). The coefficients in matrix L are defined as 1 when $i = j$, $-\alpha_{ij} / \sum_{O_{i_j} \in kNN(O_i)} \alpha_{ij}$ if

$O_{i_j} \in kNN(O_i)$ and 0 otherwise. The value of α_{ij} can be the inverse of the distance between data objects O_i and O_j, or simply equal to one.

To ensure the system has a unique solution, some constraints are imposed on the choice of control points: $\sqrt{n_i}$ control points are extracted from each cluster C_i, with $|C_i| = n_i$, and projected in the visual space employing a precise (and usually costly) method. Their projected coordinates provide the required constraints for the system.

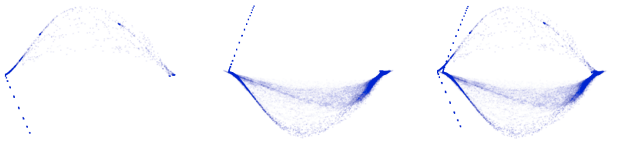
PLP has cost $O(S+C+P+I)$, where S is the complexity of the initial sampling, C is the complexity of creating the clusters, P is the complexity of projecting the control points, and I is the complexity of projecting each individual cluster. Sampling is $O(n)$, since a subset is chosen randomly from the original data. Bisecting k-means, projection of the control points, and projection of the clusters are each $O(n\sqrt{n})$. Therefore PLP has complexity $O(n + 3 * n\sqrt{n}) = O(n\sqrt{n})$.

IV. New Methods for Time Coherent Projections

In this section we describe how we have modified Fastmap and PLP to preserve temporal coherence, thus enabling their application to time-varying data.

A. TC-Fastmap - *Time Coherent Fastmap*

As described in Section III-A, Fastmap requires solving Equation 1 to obtain a projected coordinate x_i for each object, given a pair of pivots O_a and O_b. Applying the pivot selection heuristic to time-varying data will output very different pivot pairs at subsequent time steps, introducing some unexpected variations over time, as illustrated in Figure 1. The visualizations shown refer to the *IEEE Visualization 2008* data, further detailed in Section V.



(a) time step 30. (b) time step 120. (c) A single view of both time steps.

Fig. 1. Fastmap layouts relative to two distinct time steps of the (*VisContest 2008*) simulation data. The layout obtained for time step 120 is rotated as compared with that of time step 30. The rightmost figure highlights this limitation of current projection techniques.

Pivot selection should thus consider the time domain. However, a solution that requires traversing the data several times over the whole temporal domain is not feasible at interactive rates. We thus select pivots from a representative sub-set taken across multiple time steps. Tests indicated that the best representative pivots were extracted from the sub-set containing the original points of the first, final and middle time steps, applying the same heuristic as the original formulation. This approach relies on the assumption that spatial and temporal coherence hold in time-varying data, i.e., no sudden large variations occur within a small time window. This is a valid assumption when handling time-varying volumes resulting from a simulation of a single phenomenon at a suitable temporal resolution.

Moreover, computing coordinate x_i for each data object does not require traversing the data multiple times. Since pivot selection is independent of the projection step, running more iterations of the heuristic may result in a more representative pairs of pivots.

Once the pivots have been obtained, each data object can be projected separately with an iterative implementation of the *Cosine Law*, which calculates the values of a vector $V' = (V'_1, V'_2)$ where V'_d is obtained by:

$$V'_d = \frac{(d\_O_a\_V)^2 + (d\_O_a\_O_b)^2 - (d\_O_b\_V)^2}{2d\_O_a\_O_b}$$

and the value of $dist\_O_i\_O_j$ comes from:

$$d\_O_i\_O_j = \begin{cases} d(O_i, O_j) & \text{if } d = 1 \\ \sqrt{|d(O_i, O_j)^2 - (O'_{i_{d-1}} - O'_{j_{d-1}})^2|} & \text{otherwise} \end{cases}$$

B. TC-PLP- *Time Coherent Part-linear Projection*

In Section II we singled out *PLMP* [16] as the projection technique capable of preserving temporal coherence. However, it does have some limitations when applied to time-varying volumes, some shared by *TC-Fastmap* . In particular, both techniques produce layouts with data points smoothly distributed (i.e., poor visual clustering) and both achieve better results on low-dimensional data. The first limitation will cause multiple groups of similar data objects to appear mixed in the visualization, which hinders interactive exploration. The second limitation may prevent effective usage on higher-dimensional data sets, for example, when handling volumes with many derived attributes.

Seeking an alternative we have modified PLP (see Section III-B) to handle time-varying volumetric data. The underlying rationale is to recursively apply the idea of identifying small data clusters based on neighborhoods defined in the original space, embracing all time steps at once. We define the concept of "local control points" and "temporal control points", which are, respectively, the data cluster representatives at an arbitrary time step and the data cluster representatives that span multiple time steps.

The approach has been called *TC-PLP* and comprises two major steps: *preprocessing* and *projection*. In the preprocessing step, each time-stamped data volume is clustered separately, in order to obtain the local control points that will determine the projected coordinates of all the remaining data objects in this particular volume. Then, the complete set of all local control points found is clustered to identify the temporal control points. Local and temporal control points are chosen from their original clusters by selecting the $\sqrt{size(c_i)}$ data objects closest to the cluster centroid, with $size(c_i)$ being the cluster size.

As shown in Procedure 1, at each recursion step a volume is partitioned into balanced clusters of similar data objects. Then, the set of all temporal control points is projected with Fastmap (or another projection technique). In the projection

step, a neighborhood graph is built and a Laplacian linear system is solved for each recursion, first to find the projected coordinates of the local control points and then to compute the projections relative to each time stamp.

Procedure 1 Algorithm for extracting control points.

Input: D: Time-varying data set.
Output: Control points.

1) For each time-stamped volume D_t:
   a) apply the *bisecting k-means* on volume D_t to obtain clusters C_{ti}.
   b) For each cluster C_{ti}:
      i) create the kNN graph for C_{ti}.
      ii) $CP_{ti} \leftarrow$ select control points from C_{ti}.
   c) save the neighborhood information for later and return the control points in CP_{ti}.

V. Data sets

We illustrate how the proposed techniques may assist exploration of time-varying volumes on three data sets describing multivariate simulations of 3D spatial phenomena, made made available for editions of the *IEEE Visualization Contest*.

Hurricane Isabel

The *VisContest 2004* [18] data refers to a simulation of the Hurricane Isabel by the *National Center for Atmospheric Research*. $500 \times 500 \times 100$ regular grid volumes describe the behavior of several variables along 48 time steps. Volumes have been reduced with regular sampling to $125 \times 125 \times 25$. Each *voxel* is associated with 10 scalar variables and one velocity vector. 6 scalars are values of primary mixing ratios, 2 are derived mixing ratios, 2 are values for pressure and temperature. The magnitude of the velocity vector has been considered, resulting in a voxel feature vector with 11 scalar attributes.

Star explosion: ionization front

This data has been released for the *VisContest 2008* [19] and consists of multifield scalar volumes output by a simulation designed to understand the formation of galaxies, particularly the effect of *shadow instabilities* where radiation ionization fronts scatter around primordial gas. Data is organized in $600 \times 248 \times 248$ regular meshes over 200 time steps. Each cell describes 10 scalar values: total particle density, gas temperature, and 8 types of mass abundances (H, H+, He, He+, He++, H-, H2, H2+), plus a velocity vector, from which additional scalar values may be extracted, namely magnitude, divergence and curl magnitude.

We downsampled each volume by a factor of 4 by randomly selecting one cell from each group of $4 \times 4 \times 4$ cells, generating volumes with $576,600$ voxels. The temporal series was also reduced to 21 steps, corresponding to time stamps 0, 10, 20, 30, ... 170, 180, 190, 199).

Centrifugal pump

This data, released for the *VisContest 2011*, is from a simulation of a centrifugal pump described by three models, namely RANS, LES and a Hybrid LES/RANS [20]. Each volume describes one full rotation of the pump, and the simulation covers 80 time steps. Each volume is represented as an irregular volume mesh with 6 associated attributes per cell. The scalar variables are pressure (static pressure), total pressure (static pressure plus kinetic energy of the relative velocity in pressure units), total pressure in stn frame and turbulence kinetic energy (static pressure plus kinetic energy of the absolute velocity); the vector variables are velocity (in the relative system) and velocity in stn frame (the absolute system). The meshes actually describe several parts, of which we focused specifically on the rotor.

We computed $22,478$ streamlines from an equal number of random cells, and extracted feature vectors by applying the Fourier Transform to the streamlines and taking relevant FT coefficients. The resulting data comprises 39 time-stamped volumes, each with $22,478$ streamlines. Each streamline is described by a vector with 57 features; details of data preprocessing may be found elsewhere [13].

VI. Results and Evaluation

We now illustrate the applicability of the proposed techniques to time-varying volumes, assessing the layouts regarding projection time, layout quality, as measured by stress and as perceived by a human observer, and support to high dimensionality.

Stress is a measure of the layout's capability of preserving the dissimilarity between data objects, taking as reference the original data space, and is usually adopted to objectively compare layouts generated by MP techniques. It takes values in the range $[0, 1]$, with lower values indicating better preservation of dissimilarities. We shall see later that the perceived layout quality does not necessarily bear a direct relationship with the stress value: layouts with high stress happen to prove useful, e.g., to assist identification of internal structures in volumes. The "support to high dimensionality" refers to the technique's capability of producing good quality layouts even when the input feature vectors have many dimensions.

In the following we describe the results of this evaluation conducted with the projections capable of preserving temporal coherence, namely *PLMP*, *TC-Fastmap* and *TC-PLP*.

A. Computational Times

Table I details the time to project a single volume (in seconds) by *TC-Fastmap*, *TC-PLP*, *PLMP*, Fastmap and *PLP*: we have arbitrarily picked time steps 20, 90 and 01 for the Isabel, Explosion and Pump data sets, respectively. Preprocessing times are also shown when applicable – notice that preprocessing is not required for each projection, rather it is executed once for the complete series.

We compare Fastmap with a sequential implementation of *TC-Fastmap* and a parallel CPU implementation, identified in the Table as m*TC-Fastmap*. Best results were obtained on the

	Isabel		*Explosion*		*Pump*	
Technique	$pre(t)$	$proj(t)$	$pre(t)$	$proj(t)$	$pre(t)$	$proj(t)$
PLMP	0.58	0.296	0.812	0.064	0.041	0.014
Fastmap		1.926		1.488		0.194
TC-Fastmap	1.691	0.327	1.387	0.395	0.177	0.043
m*TC-Fastmap*	1.589	0.13	1.457	0.176	0.190	0.027
PLP		56.13		301.84		3.503
TC-PLP	12.575	13.205	29.272	220.18	18.037	12.153

TABLE I
EXECUTION TIMES OF THE MP TECHNIQUES, IN SECONDS. $pre(t)$ AND $proj(t)$ REFER TO PRE-PROCESSING AND TO PROJECTION TIMES, RESPECTIVELY (FASTMAP AND *PLP* REQUIRE NO PRE-PROCESSING). M*TC-Fastmap* REFERS TO A CPU PARALLEL IMPLEMENTATION OF *TC-Fastmap*.

Data set	*TC-Fastmap*	*TC-PLP*	*PLMP*
Isabel (t=12)	$3,93E-09$	0, 016490323	0, 3453625
Isabel (t=24)	$4,40E-09$	0, 011871725	0, 40194863
Isabel (t=36)	$5,62E-09$	0, 014642373	0, 3356113
Explosion (t=40)	0, 004308424	0, 007838739	0, 007070283
Explosion (t=100)	0, 005870595	0, 017687509	0, 013870152
Explosion (t=160)	0, 009910159	0, 020783341	0, 018278432
Pump (t=8)	0, 5842294	0, 36870223	0, 598145
Pump (t=15)	0, 60382056	0, 35293133	0, 67759186
Pump (t=22)	0, 56639427	0, 29616678	0, 50004995

TABLE II
STRESS VALUES OF *TC-Fastmap*, *TC-PLP* AND *PLMP* LAYOUTS.

Isabel data. Considering projection times, the sequential and parallel implementations of *TC-Fastmap* executed nearly 6 and nearly 15 times faster, respectively, than the original Fastmap. The table also shows a speed gain of nearly 4 of *TC-PLP* over *PLP*.

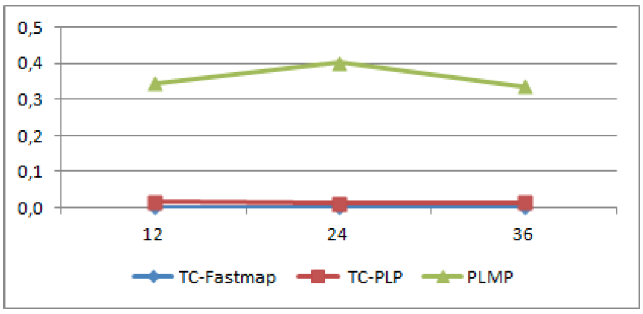
Comparing the novel approaches with *PLMP* one observes that *TC-Fastmap* executed twice as fast on the Isabel data, although in general *PLMP* was much faster considering overall (preprocessing + projection) times. Finally, *PLP* and *TC-PLP* are very slow on the Explosion data, preventing their application in interactive exploratory scenarios.

B. Layout quality: stress

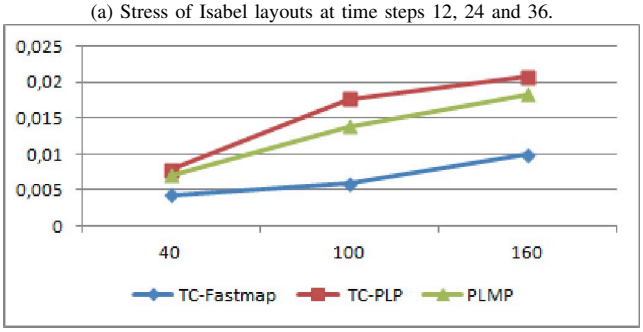
The results regarding stress are summarized in Figure 2 and Table II. Stress values for layouts obtained with *TC-Fastmap*, *TC-PLP* and *PLMP* were computed relative to three distinct time steps for each data set considered.

Figure 2a shows the results for the Isabel data: stress values of *TC-Fastmap* and *TC-PLP* are very close to 0, whereas those of *PLMP* are higher, close to 0.4. For the Explosion data, one observes in Figure 2b that *TC-Fastmap* achieved the best stress values, followed by *PLMP* and finally *TC-PLP*. The differences have not been found to be statistically significant, however. Finally, Figure 2c depicts the stress values of the Pump layouts. The layouts by *TC-PLP* have the lowest stress values, followed by *TC-Fastmap* and *PLMP*. The good results of *TC-PLP* on the Pump data set, which has higher dimensionality than the other two, suggest that it is more robust to higher dimensionalities than the other two techniques

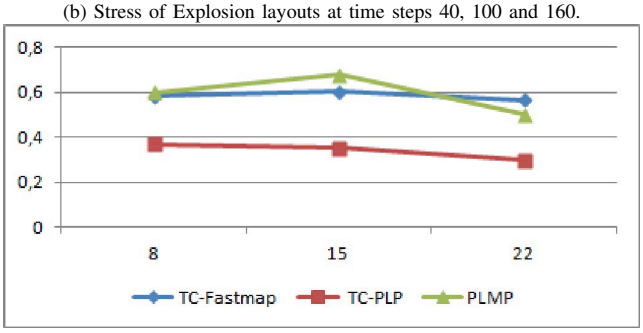
Observing the stress values in Table II one notices that with regards to stress the new techniques are comparable with *PLMP*, having resulted in similar or lower values.



(a) Stress of Isabel layouts at time steps 12, 24 and 36.



(b) Stress of Explosion layouts at time steps 40, 100 and 160.



(c) Stress of Pump layouts at time steps 8, 15 and 22.

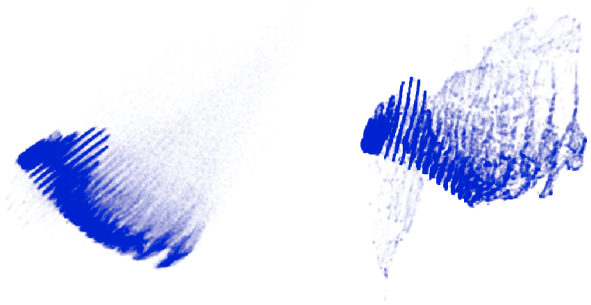
Fig. 2. Stress values of Fastmap, *TC-Fastmap*, *TC-PLP* and *PLMP* layouts, on the three data sets at distinct time steps.

C. Analyzing the 2D layouts

Let us now compare the 2D layouts obtained from the data sets with the proposed techniques with those obtained with *PLMP*, taken as the baseline.

Figures 3, 5 and 7 depict layouts obtained *TC-Fastmap* or *TC-PLP* for the Isabel, Star Explosion and Pump data sets, respectively, and the corresponding *PLMP* layout. Notice that *PLMP* and *TC-Fastmap* layouts are generally smoother than those of *TC-PLP*, i.e., data points appear more evenly distributed spatially. Because it attempts to preserve neighborhoods, *TC-PLP* favors the formation of clusters in the projected space.

Figure 3 shows layouts of the Isabel data obtained with *PLMP* and *TC-PLP*: The latter reveals small clusters that can be selected for further exploration. The *TC-PLP* layout shall be considered to further illustrate the exploratory capabilities afforded by projections. Figure 4 depicts views of a particular volume of the Isabel data. Two user selections are shown in the feature space view given by the layout, indicated by the



(a) Isabel: *PLMP* (b) Isabel: *TC-PLP*

Fig. 3. *PLMP* and *TC-PLP* layouts of the Isabel data (volume 20).

red and yellow curves. A volume rendering of the voxels in the corresponding red and yellow selections is shown at the top, revealing internal structures in the hurricane – the yellow region corresponds to the center of the storm. The thick red and orange lines indicate the two major spatial directions, and inspection shows that the placement of the feature vectors in the layout reflects the height of the spatial structures and their distance to the storm center. By mapping structure height to color intensity in the red line, and distance to the center to color intensity in the orange line, one observes increasing color intensities along both lines.

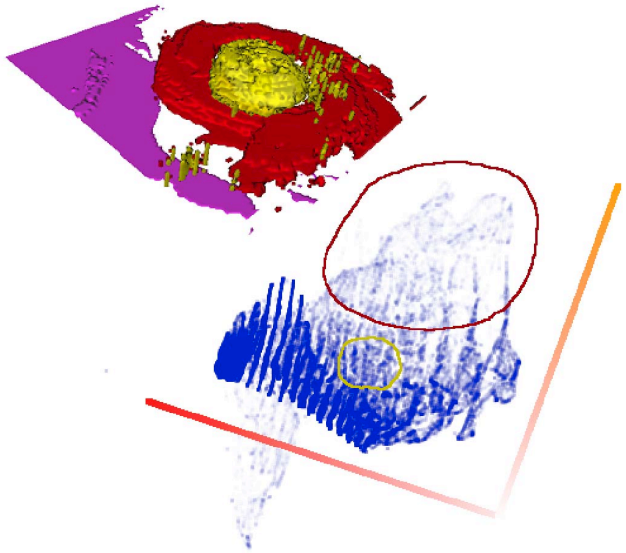
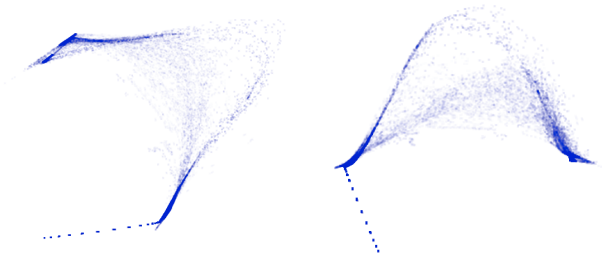


Fig. 4. Feature space projected layout of a volume from the Isabel data computed with *TC-PLP* (bottom). The user selections delimited by the yellow and red borders correspond to internal structures in the hurricane, for which an object space view is shown at the top.

For the Explosion data, both techniques compute very similar layouts for the volume at time step 90, shown in Figure 5. Further inspecting the layout by *TC-Fastmap*, Figure 6 illustrates selections of six clusters that actually correspond to distinct spatial structures. For each selection a surface rendering view of the corresponding voxels is shown in the



(a) Explosion: *PLMP* (b) Explosion: *TC-Fastmap*

Fig. 5. *PLMP* and *TC-Fastmap* layouts of the Star Explosion data (volume 90).

figure – except for the selection indicated by the black border, which corresponds to the background voxels. Again, this example illustrates that selections in the feature space allow identifying meaningful spatial structures.

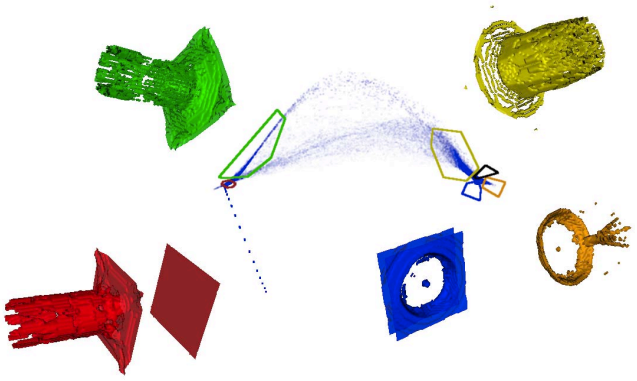
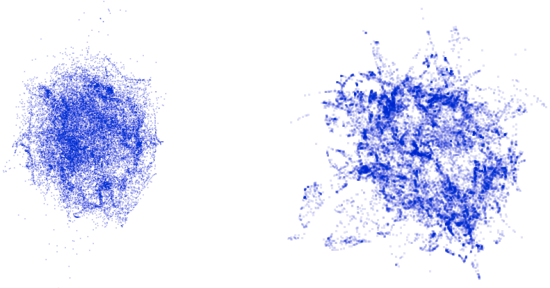


Fig. 6. Selecting groups in the *TC-Fastmap* layout of the Star Explosion data and observing the corresponding object space views.

TC-Fastmap and *PLMP* produced similar results on the Pump data, as illustrated in Figure 7 by the layouts relative to time step 01. The *TC-PLP* layout shows a spatial distribution of the feature vectors in which data points are more clustered.

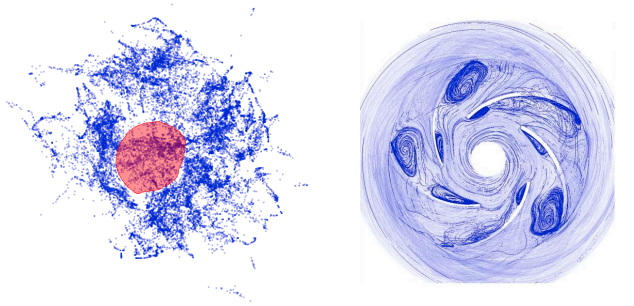


(a) Pump: *PLMP* (b) Pump: *TC-PLP*

Fig. 7. *PLMP* and *TC-PLP* layouts of the Pump data (volume 01).

Figure 8 shows the feature space view given by the layout

and a spatial view: 8a depicts the projection of the streamline features, with a user selection indicated, for which the corresponding streamlines are shown highlighted in Figure 8b. One observes the selected streamlines describe vortical structures in the flow.



(a) *streamline* projection view with a selection highlighted.

(b) *streamline* spatial view highlighting vortical structures selected in (a).

Fig. 8. *TC-PLP* projected layout of the streamline feature space of the Pump data and spatial view of the streamlines.

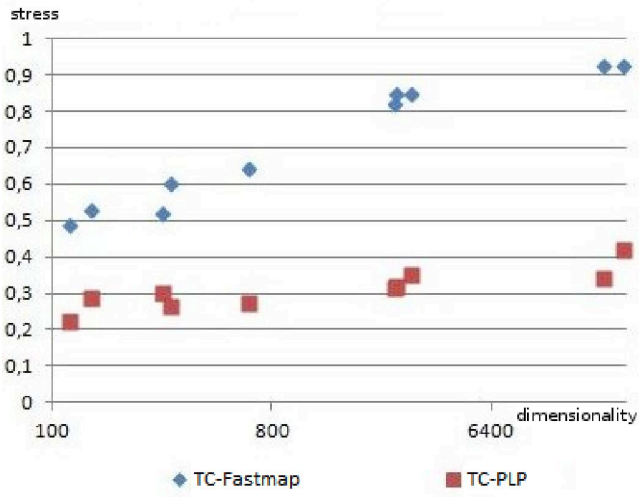


Fig. 9. Comparing stress values of *TC-Fastmap* and *TC-PLP* layouts as a function of data dimensionality.

D. High-dimensionality

We now investigate the quality of layouts generated with *TC-PLP* and *TC-Fastmap* as a function of data dimensionality. Since none of the previous data sets is characterized by very high dimensionality, for this investigation we took a textual corpus of 574 scientific articles (only title, abstract, and references considered). We generated multiple 'bag-of-words' vector representations [21] for this corpus, varying the number of 'words' (features) to obtain data sets with the following dimensionalities: 118, 144, 283, 307, 641, 2,548, 2,591, 3,000, 18,694 and 22,313. Notice that representations with 600 or more features characterize very sparse data spaces. Figure 9 shows the stress [22] values computed for *TC-PLP*and *TC-Fastmap* (PLMP cannot handle data with a point per attribute ratio above one). Stress values of *TC-PLP* are consistently smaller and increase slowly as data dimensionality increases, as compared to *TC-Fastmap*.

The previous discussion confirmed that the new techniques are suitable for handling volumetric data sets, and moreover *TC-PLP* has been shown robust to situations that require handling high-dimensional data. This, along with the support for temporal variation renders them as possible alternatives for MP-based visualization environments devised to assist exploratory investigation of simulation data.

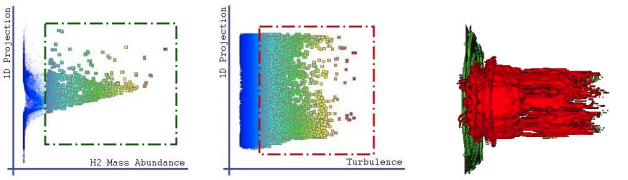
VII. *Scatter Projection*

Similarity-based MP layouts can be valuable tools for visual exploration, as users may interact to identify and analyze groups of similar data elements. However, they fail to reveal relationships amongst data attributes, or simulation variables. We introduce a hybrid visualization to support investigating the role of specific variables, which combines the familiarity of scatterplots with the segregation capability of projections.

This visualization adopts a 2D Cartesian coordinate layout, with the horizontal axis mapping a user-defined focus data variable (attribute) and the vertical axis mapping a 1-D projection of a user-selected sub-set of features. This projection may be obtained, e.g., with Fastmap on a single iteration. The resulting visualization allows focusing on the global similarity between data elements while simultaneously investigating how a target variable affects global behavior, as illustrated next.

We show how the *Scatter Projection* helps answering a question about the star explosion simulation posed at the 2008 Visualization Contest, namely: *H_2 enables primeval gas clouds to collapse and form the first stars before galaxies later coalesce. Where is H_2 most prevalent in the simulation?*

Two *Scatter Projection* views of a particular time step of the simulation are shown in Figure 10. We picked time step 99, so that findings would be comparable with reports from the literature on this problem [23]. In Figure 10a the horizontal axis maps the target attribute *H_2 Mass Abundance* and the vertical axis maps a 1D projection obtained considering a two-dimensional feature space of *density* and *temperature*. Cold/warm glyph colors map low/high values of the target variable. In the projection axis one observes that higher values of H_2 are concentrated in the middle region. In Figure 10b *turbulence* is the target in the horizontal axis, whereas the vertical axis again maps a projection of *density* and *temperature*. Unlike H_2, higher values of turbulence are not concentrated in a particular region of the projection, reflecting the prevalence of H_2 at this moment in the simulation. The view 10c shows a volume rendering of the voxels corresponding to the two selections delimited by the red and green borders in the *Scatter Projections*. Notice that the higher values of H_2 in Figure 10a appear in the volume view concentrated in the central and back regions of the advancing ionization front. The volume view reveals that high values of *turbulence* occur in the middle of the two regions depicting high values of H_2. Linsen et al. [23] discuss the prevalence of H_2 and report similar results.



(a) *Scatter Projection* focused on H_2 Mass Abundance. (b) *Scatter Projection* focused on Turbulence. (c) Object space view of selections in (a) and (b).

Fig. 10. Scatter projection visualizations of H_2 Mass Abundance (a) and Turbulence (b) versus 1D projection of *density* and *temperature*, and (c) object space volume rendering of selections indicated. (Star Explosion).

VIII. Concluding Remarks

We have reformulated Fastmap and PLP multidimensional projection techniques, introducing the *TC-Fastmap* and *TC-PLP* as projection techniques with enhanced support to volume exploration and feature space investigation of time varying volume data. They preserve temporal coherence across all time stamps. *TC-Fastmap* incorporates a pre-processing step in which pivots are selected from a representative data sample that considers the whole temporal domain. *TC-PLP* also enables selecting representative control points over the temporal domain, in this case adopting a hierarchical strategy to identify relevant control points. These techniques have been shown to produce time-coherent projections while achieving *stress* values that are comparable to or better than those of existing time-varying projections. The new techniques produce layouts with improved group separability, which helps to identify spatial regions in which voxels share similar properties, and in particular *TC-PLP* has been shown to be robust to handling high dimensional data. Additionally, the *Scatter Projection* hybrid visualization supports investigating the effect of specific data attributes on data behavior in this scenario. Further validation of the techniques is required, and we are also interested in investigating their application to further assist attribute related exploratory investigation of temporal data, including hierarchical visual attribute selection.

Acknowledgments

The authors acknowledge the financial support of FAPESP, CNPq and CAPES-PROBRAL.

References

[1] T. T. Elvins, "A survey of algorithms for volume visualization," *SIGGRAPH Comput. Graph.*, vol. 26, pp. 194–201, 1992.

[2] Y. Gu and C. Wang, "Transgraph: hierarchical exploration of transition relationships in time-varying volumetric data," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, pp. 2015–2024, 2011.

[3] C. Faloutsos and K.-I. Lin, "Fastmap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets," in *ACM SIGMOD Int. Conf. on Management of Data*. New York, NY, USA: ACM, 1995, pp. 163–174.

[4] F. V. Paulovich, D. M. Eler, J. Poco, C. P. Botha, R. Minghim, and L. G. Nonato, "Piecewise laplacian-based projection for interactive data exploration and organization," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 1091–1100, 2011.

[5] W. Chen, Z. Ding, S. Zhang, A. MacKay-Brandt, S. Correia, H. Qu, J. A. Crow, D. F. Tate, Z. Yan, and Q. Peng, "A novel interface for interactive exploration of dti fibers," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, pp. 1433–1440, 2009.

[6] G. Young and Householder, "Discussion of a set of points in terms of their mutual distances," in *Psychometrika*, vol. 3, 1938, pp. 19–22.

[7] R. Jianu, C. Demiralp, and D. Laidlaw, "Exploring 3d dti fiber tracts with linked 2d representations," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, pp. 1449–1456, 2009.

[8] P. A. Eades, "A heuristic for graph drawing," in *Congressus Numerantium*, vol. 42, 1984, pp. 149–160.

[9] M. Chalmers, "A linear iteration time layout algorithm for visualising high-dimensional data," in *IEEE Visualization*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1996, pp. 127–ff.

[10] J. Poco, D. M. Eler, F. V. Paulovich, and R. Minghim, "Employing 2d projections for fast visual exploration of large fiber tracking data," *Comput. Graph. Forum*, vol. 31, no. 3pt2, pp. 1075–1084, 2012.

[11] P. Joia, F. Paulovich, D. Coimbra, J. Cuminato, and L. Nonato, "Local affine multidimensional projection," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2563 –2571, 2011.

[12] J. D. II, E. W. Anderson, L. G. Nonato, and C. T. Silva, "Interactive vector field feature identification," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1560–1568, 2010.

[13] T. S. R. Santos, C. Wong, M. C. F. de Oliveira, R. Minghim, and D. M. Eler, "Vortices identification based on projection of streamlines," 2011, iEEE. Visualization Contest. Mini-paper contribution.

[14] H. Guo, H. Xiao, and X. Yuan, "Multi-dimensional transfer function design based on flexible dimension projection embedded in parallel coordinates," in *IEEE Pacific Visualization Symp.*, 2011, pp. 19 –26.

[15] H. Guo, H. Xiao, M. Lu, and X. Yuan, "Scalable multivariate volume visualization and analysis," in *IEEE Symp. Large Data Analysis and Visualization*, 2011, pp. 119 –120.

[16] F. V. Paulovich, C. T. Silva, and L. G. Nonato, "Two-phase mapping for projecting massive data sets," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, pp. 1281–1290, 2010.

[17] F. V. Paulovich, L. G. Nonato, R. Minghim, and H. Levkowitz, "Least square projection: a fast high-precision multidimensional projection technique and its application to document mapping," *IEEE Trans. Visualization and Computer Graphics*, vol. 14, pp. 564–575, 2008.

[18] W. Wang, C. Bruyere, and B. Kuo, "Competition data set and description," 2004, in IEEE Visualization Design Contest.

[19] D. Whalen and M. L. Norman, "Competition data set and description," 2008, in IEEE Visualization Design Contest.

[20] G. D. W.-I. A. L. Institute of Applied Mechanics, Clausthal University, "Competition data set and description," 2011, iEEE Visualization Design Contest.

[21] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," in *Information Processing and Management*, 1988, pp. 513–523.

[22] S. Ingram, T. Munzner, and M. Olano, "Glimmer: multilevel mds on the gpu," *IEEE Trans. Visualization and Computer Graphics*, vol. 15, no. 2, pp. 249–261, 2009.

[23] L. Linsen, T. V. Long, and P. Rosenthal, "Linking multi-dimensional feature space cluster visualization to surface extraction from multi-field volume data," *IEEE Computer Graphics and Applications*, vol. 29, no. 3, pp. 85–89, 2009.