EuroVis Workshop on Visual Analytics (2013)
M. Pohl and H. Schumann (Editors)

Feature-Based Visual Analytics for Studying Simulations of Dynamic Bi-Stable Spatial Systems

C. Eichner [1] A. Bittig [2] H. Schumann [1] C. Tominski [1]

[1]Computer Graphics [2]Simulation and Modeling, Institute for Computer Science, University of Rostock, Germany

Abstract

Simulations of dynamic bi-stable spatial systems usually generate large and complex data that are hard to evaluate. In this paper, we describe how visual analytics technology can help in analyzing such simulation data. The idea behind our approach is to utilize concepts of feature-based visualization. Consequently, we consider (1) interactive specification of meaningful features, (2) analytic extraction and tracking of features as well as detection of events in the features' evolution, and (3) visual representation of features with their spatial, temporal, and structural aspects. Our solution has been used by simulation experts to analyze spatio-temporal distributions of multiple types of particles in reaction-diffusion simulation data. With the help of the feature-based approach the scientists were able to understand how the spatial separation of proteins develops over time.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Miscellaneous—Visualization of simulation data

1. Introduction

Computer simulations of biochemical systems are a powerful means to develop an understanding of natural phenomena. In contrast to real-life observations, simulations usually provide a more cost-effective and easier way to get data of the phenomena under investigation. However, the generated data are usually large and complex making it necessary to provide appropriate tools for their analysis.

Previous work suggests that interactive visual approaches are useful for supporting the analysis of simulation data [Dol07, US09]. However, plainly following Tufte's "*Above all else show the data.*" will not suffice when the data are larger. In such cases it is necessary to provide tools that enable the user to focus on relevant and digestible subsets of the data.

A classic approach with exactly the rationale to focus on meaningful parts of the data is feature-based visualization [RPS01]. Based on a specification of what is relevant, features are automatically extracted from the data and tracked over time. Higher-level events in the evolution of features are detected as well. The visual representations show features and events, rather than the underlying raw data. The user can concentrate on the information that is important for the task at hand; less-relevant data are omitted.

In this work, we utilize concepts of feature-based visualization in order to support the analysis and exploration of larger simulation data. While there is previous work in visual analytics that incorporates the one or the other aspect of the feature-based approach (usually feature specification and extraction), the aspects that lead to higher-level insight (i.e., feature tracking and event detection) are considered only rarely, if at all. Our solution realizes a complete pass through the pipeline of the feature-based strategy as suggested in [RPS01]. Where necessary we adapt the classic methods to meet the requirements of our users.

Next we will take a closer look at this scenario. In Section 3 we will describe the internals of our feature-based approach. A brief discussion concludes our work in Section 4.

2. Background and Related Work

Our studies have been conducted in collaboration with simulation experts who need effective tools to support the investigation of simulations of dynamic bi-stable spatial systems. The researchers simulate such systems using the Next Subvolume method [EE04] with ML-Space [BHMU11]. The model contains two substrate proteins produced by one of two corresponding enzymes. Each enzyme blocks the production capability of the respective other by binding to it.

DOI: 10.2312/PE.EuroVAST.EuroVA13.025-029

The simulation uses mesoscopic methods to model particle distribution and movement. The 3D simulation space is partitioned into sufficiently small *subvolumes* such that a homogeneous particle distribution can be assumed for each subvolume. As the simulation progresses, particles diffuse between subvolumes, and particles partake in reactions (i.e., binding or producing) within subvolumes.

The data generated by such simulations contain information about the spatial movement of thousand of microscopic particles and their interaction with each other. The simulation experts are interested in identifying three dimensional spatial separations of the different types of proteins at given points in time and in understanding the general dynamic development of the spatial distribution of proteins over time.

Previous work by Unger et al. [US09] utilizes multiple coordinated views and direct volume rendering to visualize proteins' spatial distribution. Luboschik et al. [LTB*12] focus on visualizing trajectories of simulated particles. However, as these low-level methods basically show each and every detail of the data they reach their limits when it comes to identifying and evaluating key characteristics. What is needed are higher-level visualizations that focus on giving a spatio-temporal overview of core features in the simulated bio-chemical systems.

About a decade ago Reinders et al. [RPS01] formulated the theoretical foundations behind an approach that is able to generate higher-level overviews of large and complex data sets by reducing the shown information to relevant features. While Reinders et al. describe a general feature-based visualization pipeline, concrete applications of it are almost exclusively related to the visualization of flow data.

A recent survey by Kehrer and Hauser [KH13] indicates the potential of the feature-based concept. For example, Anand et al. [ADW12] compute features to guide exploration of multivariate data, Wong et al. [WFA*03] extract features to speed up computations, and Kandogan [Kan12] works with annotations derived from statistical analysis. But often the focus is on feature specification and extraction and not so much on feature tracking and event detection. An example that includes tracking is the work by Rohrdantz et al. [RHD*12], who analyze text document streams.

Here we aim to apply feature-based concepts in the context of visual analytics of simulation data generated by the Next Subvolume method. In our scenario we need the entire feature-based pipeline, because the evolution of features over time and events in the features' evolution are of primary interest to the simulation experts. To this end, we adapt the classic feature-based approach and extend it where needed.

3. General Approach

The next paragraphs will describe in more detail how we realized the complete feature-based pipeline to handle spatial and temporal aspects of the simulation data:

- We show how a meaningful specification of features can be achieved in the context of our application scenario.
- We extract features from the simulation data and visualize them as 3D ellipsoids in their spatial frame of reference.
- We track features over time and detect events in their evolution. The resulting tracking graph is visualized to convey the temporal aspects of the data.
- We provide interactive tools such as coordinated selection across views and dynamic filtering to support the exploration and analysis of the data.

3.1. Feature Specification

A suitable specification is required to be able to extract meaningful subsets of the simulation data. In the area of flow visualization, the origin of the feature-based approach, plausible feature definitions exist (e.g., vortices, shock waves, critical points). In our setting, there are no such a-priori definitions of features.

Discussions with our users revealed that they are interested in features based on protein concentration and related attributes. Using such features they could determine regions where certain proteins are dominant, which in turn allows them to study distribution and spatial separation of the proteins. However, the thresholds of dominant concentrations vary depending on the simulated system. Therefore, we have to resort to an interactive and exploratory specification procedure.

We follow a practical approach that integrates spatial, temporal, and attribute aspects similar to [DGH03, GRBM11]. Basic data characteristics (i.e., the frequency distribution of protein concentration) are conveyed in parallel aligned histograms. Users can then perform brushing operations on the histograms to capture the parts of the data they deem interesting and relevant. This results in a set of logical rules to be handed over to the feature extraction.

3.2. Feature Extraction and Spatial Visualization

With the help of the logical rules defined in the specification phase, we can classify the individual subvolumes in the 3D simulation space. A subvolume may or may not match the defined rules. Neighboring subvolumes that match identical rules are merged to form coherent regions or *features*. This matching is performed for each time step individually.

In our simulation setting it is not uncommon that proteins are distributed evenly across the 3D simulation space. Under such circumstances the classic extraction process generates large numbers of tiny features. However, the simulation experts need an overview of the principle characteristics of the data in the first place. Details of fine-grained features would distract from the important changes in the distribution of the extracted features.

Therefore, we further adapt the extraction to our users'

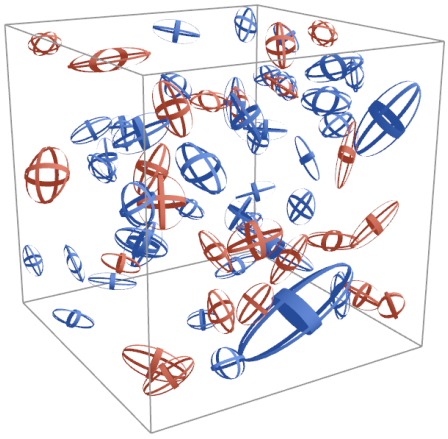

Figure 1: *Ellipsoids representing regions with high concentration of two selected proteins.*

needs. We use a combination of connectedness (e.g., shared point, edge, or face) and Euclidean distance as the criterion for merging of subvolumes to features. This way, we give the users the opportunity to steer the size of the extracted features, where they often favor larger features. Secondly, the user can explicitly filter out smaller features.

The features that pass the filter are mapped to 3D ellipsoid glyphs as suggested in [WPSP96]. Each glyph approximates spatial distribution of merged subvolumes by encoding average position, volume alignment, and size. The ellipsoids are placed in a three dimensional representation of the simulation space. Figure 1 shows an example with several features. Red and blue ellipsoids stand for regions with high concentration of two different proteins (see [EE04] for details).

The ellipsoid representation helps users in evaluating the spatial distribution of proteins at any selected point in time. Insight into to the temporal development can be gained by comparing the features of different time points. However, a semantic relationship between features across multiple time points remains hard to detect visually. This task is supported by feature tracking.

3.3. Feature Tracking and Temporal Visualization

The feature tracking step establishes relationships between features of consecutive time steps. A feature at time t_i is related to another feature at time t_{i+1}, if the latter describes the "evolved" version of the former. In addition to *one-to-one* relations (i.e., feature continues to exist), there can also be *one-to-many* or *many-to-one* relations when features split or merge, which might indicate important events in the evolution of the data [SSZC94].

Because the concentrations in our simulated systems change only slowly, the spatial properties of features in subsequent time steps are very similar. This allows us to use a region-based algorithm (e.g., [SW97]) to track features and detect events.

The result of the tracking algorithm can be interpreted as a *feature graph* (or event graph, see [RPS01, CSJ03, BWT*11]) in which nodes represent features and edges connect related features. Paths in the graph links features across multiple time steps, thus establishing a semantic relationship of features over time. Particular connectivity patterns in the graph represent events. For instance, a node with one incoming edge and multiple outgoing edges corresponds to a split event. Although the feature graph abstracts from the spatial properties of features, it yet crystallizes their temporal evolution.

We visualize the feature graph (1) in a dedicated view that focuses on temporal aspects and (2) as a novel fusion of spatial, temporal and structural aspects.

To emphasize the temporal character of the feature graph, we compute a layered drawing based on the Sugiyama layout [STT81]. Each time step is associated with a layer that contains all features extracted from that time step properly stacked along the vertical axis. Layers are arranged along the horizontal axis ordered by time. The visualization uses node size to encode feature "size" (e.g., volume, avg. concentration) and small symbols indicate events (i.e., split or merge). Connected components in the graph are colored with a distinct hue depending on the dominant type of protein.

As illustrated in Figure 2, one can easily spot bigger features, and by comparing connected nodes it is possible to identify features that grow or shrink. By comparing sizes and numbers of nodes contained in a layer it is even possible to estimate which protein occupies more space and whether there are few bigger spots of higher concentration or several smaller ones. This helps to identify whether there is a sharp separation of proteins in larger areas or a more homogeneous distribution.

But with the abstract structural representation of the tracking graph alone it obviously remains difficult to fully grasp the spatial aspects of features. Therefore, we propose to embed the graph structure directly into the 3D representation of the feature ellipsoids of a selected time interval. As illustrated in Figure 3 such a visualization conveys spatial, temporal, and structural aspects.

To find a suitable compromise to the conflict over exploring overviews with many features of many time steps or analyzing details of specific aspects for fewer features, the users must be provided with flexible interaction mechanisms.

3.4. Interaction

Our solution communicates different aspects of the same data in different views. To assist users in mentally linking the different aspects, we support interactive selection of features in any view and highlight the selected features in all views.

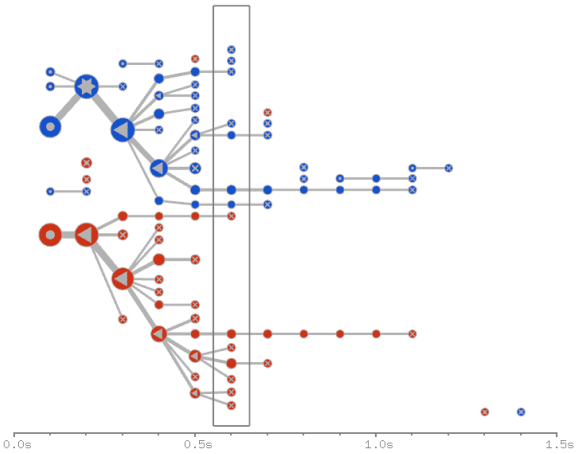

Figure 2: *Layered drawing of a tracking graph of features related to high concentration of two selected proteins.*

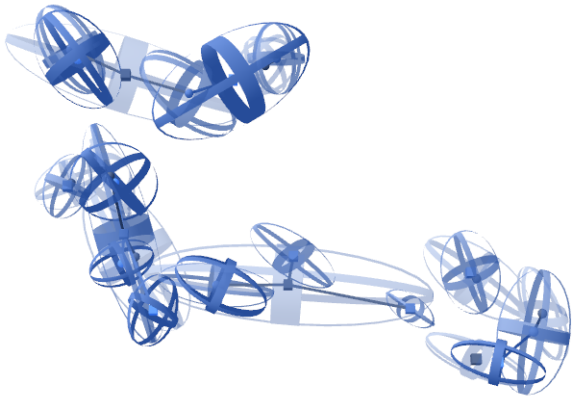

Figure 3: *Ellipsoids with embedded graph for three consecutive time steps show spatial, temporal, and structural aspects of features. Features of past time steps are dimmed.*

In addition to the standard selection of individual features, we developed novel selection modes that exploit the structure of the feature graph. The user can select (1) all features that are connected through a path in the graph, or (2) all features that belong to the same connected component. The different selection modes can be combined to form any subset of features with just a few clicks. Highlighting the selected features in all views makes it possible to evaluate features in their temporal context in one view while assessing their spatial properties in another view.

Additionally, users can restrict the selection to a specific time range (e.g., single time step, past, future, or free interval). This temporal filtering is carried out with the help of the gray frame depicted in Figure 2. To put more visual emphasis on the currently focused time step, features of past or future time steps can be dimmed or omitted altogether (see Figure 3).

4. Discussion and Conclusion

In this work, we presented visual analytics support for the investigation of dynamic bi-stable spatial systems.

The analytical component of our solution is largely inspired by the classic feature-based strategy, where it is worth mentioning that we implemented the full pipeline, including feature specification, feature extraction, feature tracking, and event detection.

In terms of visual components, we built on well-accepted representations and improved them in details. Spatial aspects of the data are visualized by 3D ellipsoids. Temporal aspects are dealt with by showing (optionally dimmed) ellipsoids of multiple time steps, and by using a dedicated layout of the feature graph. This layout also communicates the higher-level structural aspects and corresponding events of the temporal evolution of features. The embedding of the feature graph into the 3D ellipsoid display establishes a direct connection between space, time, and structure.

Interaction facilities enable the users to flexibly specify the features they are interested in and to fine-tune the extraction process with additional filter thresholds. The user can select and highlight features in a coordinated way and focus investigations on specific points or intervals in time.

The developed feature-based visual analytics approach has been used by simulation experts to observe the spatial distribution of different types of proteins in a three dimensional space, as well as to analyze the temporal changes of those distributions. By using different types of feature definitions for each observed protein, the major distribution of particles and the general dynamical behavior could be observed. For example, when features are getting smaller over time (as shown in Figure 2), there are not as many areas with high protein concentrations, leading to the impression that the two types of proteins mix up. This general view helped in analyzing the conditions under which proteins separate and the specific simulation parameters that influence the development of a system.

For up to $5 \cdot 10^5$ subvolumes, our tool is able to extract interesting features per time step at interactive rates. Tracking the features over 20 time steps is completed within a few seconds, depending on the number of features to be tracked.

In the future, we plan to combine our high-level feature-based visualization with suitable low-level representations of the raw data (e.g., 2D slices [US09] or protein trajectories [LTB*12]). While features are useful to identify interesting parts of the data, the low-level techniques will help to analyze the interesting parts in more detail.

References

[ADW12] ANAND A., DANG T. N., WILKINSON L.: Visual pattern discovery using random projections. In *Proceedings of the Conference on Visual Analytics Science and Technology (VAST)* (2012), IEEE Computer Society, pp. 43–52. doi:10.1109/VAST.2012.6400490. 2

[BHMU11] BITTIG A. T., HAACK F., MAUS C., UHRMACHER A. M.: Adapting rule-based model descriptions for simulating in continuous and hybrid space. In *Proceedings of the International Conference on Computational Methods in Systems Biology (CMSB)* (2011), ACM, pp. 161–170. doi:10.1145/2037509.2037533. 1

[BWT*11] BREMER P.-T., WEBER G., TIERNY J., PASCUCCI V., DAY M., BELL J.: Interactive exploration and analysis of large-scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics 17*, 9 (2011), 1307–1324. doi:10.1109/TVCG.2010.253. 3

[CSJ03] CHEN J., SILVER D., JIANG L.: The feature tree: visualizing feature tracking in distributed amr datasets. In *Proceedings of the Symposium on Parallel and Large-Data Visualization and Graphics (PVG)* (2003), IEEE Computer Society, pp. 103–110. doi:10.1109/PVGS.2003.1249048. 3

[DGH03] DOLEISCH H., GASSER M., HAUSER H.: Interactive feature specification for focus+context visualization of complex simulation data. In *Proceedings of the Symposium on Data Visualisation (VisSym)* (2003), Eurographics Association, pp. 239–248. URL: http://diglib.eg.org/EG/DL/WS/VisSym/VisSym03/239-248.pdf. 2

[Dol07] DOLEISCH H.: Simvis: interactive visual analysis of large and time-dependent 3d simulation data. In *Proceedings of the Winter Simulation Conference (WSC)* (2007), IEEE Press, pp. 712–720. doi:10.1145/1351542.1351674. 1

[EE04] ELF J., EHRENBERG M.: Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biology 1* (2004), 230–236. doi:10.1049/sb:20045021. 1, 3

[GRBM11] GRUCHALLA K., RAST M., BRADLEY E., MININNI P.: Segmentation and visualization of multivariate features using feature-local distributions. In *Proceedings of the International Symposium on Visual Computing (ISVC)* (Berlin, Heidelberg, 2011), Springer-Verlag, pp. 619–628. doi:10.1007/978-3-642-24028-7_57. 2

[Kan12] KANDOGAN E.: Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *Proceedings of the Conference on Visual Analytics Science and Technology (VAST)* (2012), IEEE Computer Society, pp. 73–82. doi:10.1109/VAST.2012.6400487. 2

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 495–513. doi:10.1109/TVCG.2012.110. 2

[LTB*12] LUBOSCHIK M., TOMINSKI C., BITTIG A. T., UHRMACHER A. M., SCHUMANN H.: Towards interactive visual analysis of microscopic-level simulation data. In *Proceedings of the Annual SIGRAD Conference, Special Theme: InteractiveVisual Analysis of Data* (2012), Linköping University Electronic Press, pp. 91–94. URL: http://www.ep.liu.se/ecp_article/index.en.aspx?issue=081;article=013. 2, 4

[RHD*12] ROHRDANTZ C., HAO M. C., DAYAL U., HAUG L.-E., KEIM D. A.: Feature-based visual sentiment analysis of text document streams. *ACM Trans. Intell. Syst. Technol. 3*, 2 (2012), 26:1–26:25. doi:10.1145/2089094.2089102. 2

[RPS01] REINDERS F., POST F. H., SPOELDER H. J.: Visualization of time-dependent data with feature tracking and event detection. *The Visual Computer 17*, 1 (2001), 55–71. doi:10.1007/PL00013399. 1, 2, 3

[SSZC94] SAMTANEY R., SILVER D., ZABUSKY N., CAO J.: Visualizing features and tracking their evolution. *Computer 27*, 7 (1994), 20–27. doi:10.1109/2.299407. 3

[STT81] SUGIYAMA K., TAGAWA S., TODA M.: Methods for visual understanding of hierarchical system structures. *Systems, Man and Cybernetics, IEEE Transactions on 11*, 2 (1981), 109–125. doi:10.1109/TSMC.1981.4308636. 3

[SW97] SILVER D., WANG X.: Tracking and visualizing turbulent 3d features. *Visualization and Computer Graphics, IEEE Transactions on 3*, 2 (1997), 129–141. doi:10.1109/2945.597796. 3

[US09] UNGER A., SCHUMANN H.: Visual support for the understanding of simulation processes. In *Proceedings of the Pacific Visualization Symposium (PacificVis)* (2009), IEEE Computer Society, pp. 57–64. doi:10.1109/PACIFICVIS.2009.4906838. 1, 2, 4

[WFA*03] WONG P. C., FOOTE H., ADAMS D., COWLEY W., THOMAS J.: Dynamic visualization of transient data streams. In *Proceedings of the Symposium on Information Visualization (InfoVis)* (2003), IEEE Computer Society, pp. 97–104. doi:10.1109/INFVIS.2003.1249014. 2

[WPSP96] WALSUM T., POST F. H., SILVER D., POST F. J.: Feature extraction and iconic visualization. *IEEE Transactions on Visualization and Computer Graphics 2*, 2 (1996), 111–119. doi:10.1109/2945.506223. 3