Analyzing Simulations of Biochemical Systems with Feature-Based Visual Analytics

Christian Eichner[a], Arne Bittig[b], Heidrun Schumann[a], Christian Tominski[a]

[a]Computer Graphics, Institute for Computer Science, University of Rostock, Germany
[b]Modelling and Simulation, Institute for Computer Science, University of Rostock, Germany

Abstract

Spatial simulations of biochemical systems are carried out to gain insight into nature's underlying mechanisms. However, such simulations are usually difficult to set up and they generate large and complex data. In order to help scientists understand their models and the data generated by the simulations, appropriate visual support can be a decisive factor. In this paper, we apply and extend ideas of feature-based visualization to develop a visual analytics approach to analyze data of reaction–diffusion system simulations. Our approach enables simulation experts to interactively specify meaningful features, which are automatically extracted and tracked via analytical means. Events in the features' evolution over time are detected as well. Features and events are visualized via dedicated 3D and 2D views, which in combination portray the interplay of the spatial, temporal, and structural aspects of the simulation data. Our approach is being implemented in the context of a multi-view multi-display visualization environment. We demonstrate how researchers can analyze spatio-temporal distributions of particles in a multi-step activation model with spatial constraints. The visual analytics approach helped to identify interesting behavior of the spatial simulation, which was previously only speculated about, and to examine and discuss competing hypotheses regarding possible reasons for the behavior.

Keywords:
Visual analytics, feature detection, feature tracking, feature visualization, simulation

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. Still the involved models and the corresponding simulations are complex and generate large and complex data. Appropriate tools for analyzing the data is a crucial aspect in this field of research.

Previous work suggests that interactive visual approaches are useful for supporting the analysis of simulation data [1, 2]. However, plainly following Tufte's "*Above all else show the data.*" [3] 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.

Feature-based visualization [4] is a classic approach with exactly the rationale to focus on meaningful parts of the data. Based on a formal specification of what is relevant, features are automatically extracted from the data and tracked over time. To structure the evolution of features, higher-level events are pinpointed in time. The visualization then shows features and events, rather than the underlying raw data. Because less-relevant data are omitted, users can concentrate on the information that is important to the task at hand.

In this work, we utilize the classic concepts of feature-based visualization in order to support the analysis of larger simula-

tion data with a biochemical background. 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 [4]. Where necessary we adapt the classic methods to meet the requirements of the biochemical simulation data and the simulation experts' needs.

Our research is based on established visualization concepts. The key contribution of this paper is to show how to combine and extend these concepts to make them applicable to new real-world challenges in simulation and modeling research. Prior to our studies, simulation experts had only few options to tackle their new data analysis problems, while the visual analytics literature had no solutions readily available that would address the requirements in that field of research sufficiently.

This paper is an extended version of a contribution [5] to the 2013 EuroVis workshop on Visual Analytics (EuroVA). The present version has been thoroughly revised and extended with more detailed descriptions of the data and task being addressed, the related work, and the general approach. As new content we included the implementation of our approach in a novel multi-view multi-display visualization environment. Further we added a detailed use case that demonstrates how simulation experts have applied the feature-based visual analytics solution to gain a better understanding of a real-world simulation problem.

Email address: `christian.eichner@uni-rostock.de` (Christian Eichner)

In the next section, we will take a closer look at the simulation scenario we aim to support. In Section 3 we will describe the internals of our feature-based approach in detail. The implementation of the approach in a multi-view multi-display visualization environment will be presented in Section 4. A detailed use case illustrating insights gained with the help of the proposed solution is part of Section 5. A brief discussion and preliminary user feedback are given in Section 6. Section 7 concludes this work.

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 reaction–diffusion systems [6]. The simulation uses mesoscopic methods to model distribution and movement of biochemical particles (e.g., proteins). Instead of simulating individual particles, an abstraction takes place reducing the model to discrete regions and the number of particles within them. To this end, the 3D simulation space is partitioned into sufficiently small subvolumes. Each subvolume stores the amounts of the different types of particles that are involved in the reaction–diffusion system. As the underlying biochemical process carries on, particles diffuse in space and partake in reactions (e.g., production, consumption, binding). In the corresponding simulation, this is reflected as an exchange of particles between subvolumes for the diffusion (i.e., absolute frequency varies) and as a change of the amount of the different types of particles within a subvolume for the reactions (i.e., relative frequency varies).

The simulation approach used here follows the Next Subvolume method [7] as realized in ML-Space [8], a model description language and simulation approach that supports particle-based and hybrid simulation. The data generated by such simulations contain information about the spatial movement of thousand of microscopic particles (e.g., different kinds of proteins) and their interaction with each other. Various aspects are relevant when simulation experts analyze the data in order to understand the simulated biochemical phenomena:

- It is important to get an overview of the spatial distribution of particles in the 3D simulation space to determine whether particles are distributed evenly or if there are any regions with significantly low or high particle concentration.

- The temporal dimension of the simulation data has to be considered to allow for the investigation of how the particle distribution is changing over time.

- The spatio-temporal development of the system needs to be analyzed with regard to the influence of different types of particles. Depending on protein types, the chemical reactions can result in spatial separation or correlation of proteins over time.

Previous work by Unger et al. [2] utilizes multiple coordinated views and direct volume rendering to visualize the spatial distribution of proteins. A 2D plane with pixel based icons

shows the mixing ratio of proteins. Gribble et al. [9] present an approach to visualize large and time-varying particle datasets using ray tracing. The visual encoding, including color mapping, shadows, and range culling can be controlled by the user at run time. Luboschik et al. [10] focus on visualizing trajectories of simulated particles. A visual grouping of particles is achieved by coloring trajectories based on the similarity in movement directions. Bürger et al. [11] and Krüger et al. [12] adapt the appearance of particle primitives, according to user defined regions of interest and particle properties. However, specification and visualization of interesting particles mainly focus on emphasizing particle flow, rather than spatial distribution of different types of particles.

However, because 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 biochemical systems.

About a decade ago Reinders et al. [4] formulated the theoretical foundations behind an approach that is able to generate higher-level overviews of large and complex time-dependent data sets by reducing the shown information to relevant *features*. Reinders et al. describe a general feature-based visualization pipeline, including feature specification, feature extraction, feature tracking, and event detection, as well as visualization of features and events.

A recent survey by Kehrer and Hauser [13] illustrates the great potential of the feature-based concept in general. Concrete applications can be found in many contexts. For example, in flow visualization, features are used to mark critical points in a vector field, to visualize the topology of the flow, or to outline similar motion trajectories [14]. Tzeng et al. [15] utilize learning algorithms for automatic feature specification and tracking in large scale flow simulations. Anand et al. [16] compute features to guide exploration of multivariate data. In this case features represent outliers, clusters of similar data points or sets of points that match a specific pattern. Kandogan [17] also works with annotations derived from statistical analysis. Jänicke et al. [18] use features to visualize high dimensional data with 2D clouds. Wong et al. [19] extract features to speed up computations in the context of analyzing data streams. The work by Rohrdantz et al. [20] incorporates feature tracking to analyze text document streams. In visual analytics of video data, feature extraction and tracking is used to recognize moving objects and determine their trajectories [21]. An example from simulation of particles can be found in Grottel et al. [22], where features are used to visualize the evolution of spatial molecule clusters. In addition to a three dimensional representation also a schematic view is provided for investigating interexchange of molecules between clusters over time.

As the previous list of existing work illustrates, the general feature-based approach can be applied in many domains. However, concrete features, their definition, meaning, and usage, highly depend on the application background. 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 visual analytics pipeline to handle spatial, temporal, and structural aspects of the simulation data:

- We show how a meaningful specification of features can be achieved in the context of multivariate spatio-temporal simulation data.

- We extract features per time step 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 temporal and structural aspects.

- We provide interactive tools to support the coordinated exploration and analysis of the data of different or differently configured models and simulators.

3.1. Feature Specification

A suitable formal specification is required to be able to computationally extract meaningful subsets of the simulation data. In the area of flow visualization, the origin of the feature-based approach, plausible feature definitions exist, for instance to extract vortices, shock waves, or critical points. In our setting, there are no such a-priori definitions of features. Therefore, we pursue an interactive visual approach to feature specification.

Discussions with our users made clear to us that they are interested in features based on the concentration of protein particles 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 are not precisely fixed and vary depending on the simulated system. Therefore, we must support a flexible and exploratory specification procedure.

We follow a practical approach similar to [23, 24]. The idea is to visualize basic data characteristics and let the user select regions of interest. Figure 1 shows the frequency distribution of the particle count of proteins as parallel aligned histograms. Users can perform brushing operations on the histograms to capture the parts of the data they deem interesting and relevant. Technically, the brushing is based on open or closed intervals of interesting value ranges and their logical combination via AND and OR operators.

The feature specification is accompanied with a basic visualization that highlights the subvolumes that match the value
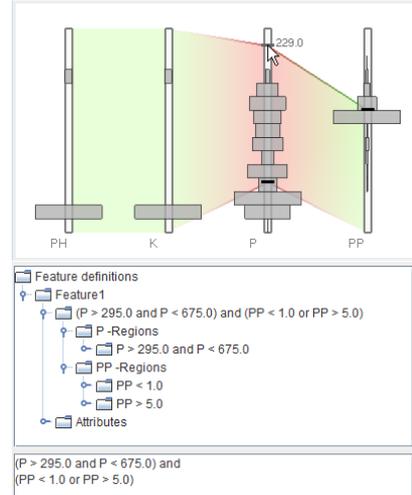

Figure 1: Histograms visualizing value distribution across multiple dimensions. Brushing intervals on the histogram axes serves to create formal logical definitions of features to be extracted. A tree view shows the structure of the feature definition and a text field can be used by advanced users for textual input.

ranges being constructed. By this we achieve a tight coupling of the interactive specification means with a visual preview of the effects of interval thresholds and logical rules.

In our application domain, regions with locally high or low particle concentrations are of particular interest. For such regions of interest, brushing on absolute particle counts alone is not fully sufficient because the local particle concentrations might vary across different regions and might be only slightly different from the global average. This makes it difficult to find good thresholds to capture all relevant subvolumes.

To address this problem, we derive additional data attributes that consider the local neighborhood in the 3D simulation space. We compute the gradient magnitude and the Laplace operator as basic means to measure local spatial changes and visualize them as additional axes in the histogram display. The derived attributes can be brushed to crystallize 3D regions with certain local variance with respect to their surrounding space.

Eventually, the specification procedure yields logic formulas, one for each type of feature (e.g., related to the different types of entities taking part in a simulation). Formulas can be stored for later reuse and fine tuning. Advanced users can also directly enter formulas if they prefer to (see Figure 1). Once all relevant feature have been specified, the constructed formulas are handed over to the feature extraction, which will be described next.

3.2. Feature Extraction and Spatial Visualization

With the help of the logic formulas defined in the specification phase, we classify the individual subvolumes in the 3D simulation space. A subvolume may or may not match a specified feature definition. In order to form coherent regions or *features*, the next step is to merge neighboring subvolumes that do match. Analogous to the classic feature-based approach, the feature extraction is performed for each type of feature and for each time step separately.
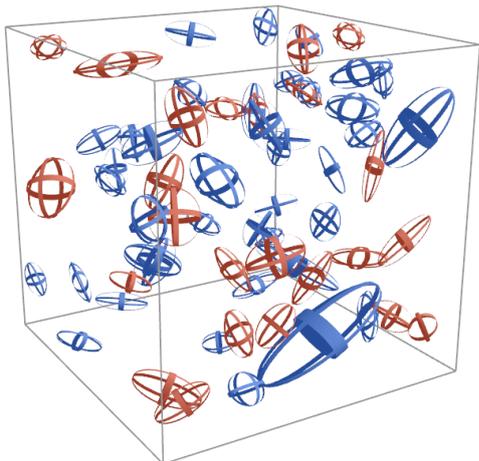
Figure 2: Ellipsoids representing regions with high concentration of two selected proteins (in a simulation of the double negative feedback system of [7]).
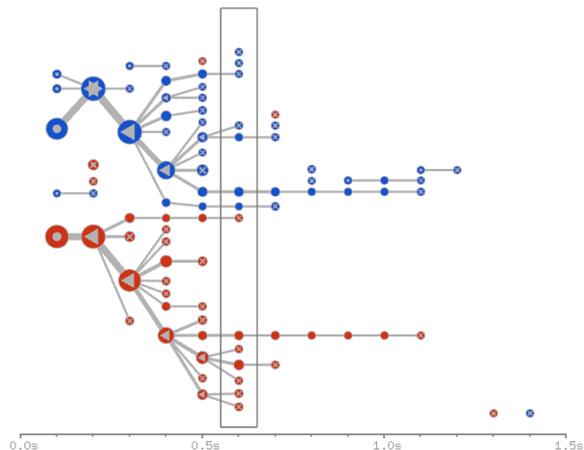

Figure 3: The tracking graph of features related to high concentration of two selected proteins shown as red and blue nodes. Nodes are horizontally arranged according to time from left to right and vertically arranged so as to minimize edge-crossings. Node size indicates feature "size" and symbols indicate events in the features' development.

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 a large number of tiny features. However, the simulation experts need an overview of the principal 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 slightly adapt the extraction procedure. To be able to control the size (and thus the number) of the features to be extracted, we use a region growing approach in combination with a flexibly adjustable Euclidean distance threshold. The distance threshold d enables us to model connectedness criteria that must hold for subvolumes to be merged to features. The following connectedness criteria are useful: subvolumes have to share a face ($d = 1$), an edge ($d = \sqrt{2}$), or a point ($d = \sqrt{3}$). Alternatively, subvolumes can be disconnected ($d > \sqrt{3}$), but their distance must be smaller than d. The distance d can be controlled interactively while visual feedback constantly indicates the extraction outcome. This way, the user can explicitly filter out smaller features until the size of the extracted features is sufficiently large.

Once features have been extracted they need to be visualized. A primary concern is to show the spatial characteristics of features. The visualization literature suggests different alternatives, including highlighting, iso-surfaces, ellipsoids, or skeletons [14]. Our users are typically interested in compact regions where certain model entities cluster or at least are over-represented relative to others. Further, it is more important to roughly identify the distribution and size of extracted features, rather than the individual subvolumes that features encompass. Under these circumstances 3D ellipsoids as suggested in [25] appear to be a good choice to form an understanding of the underlying processes of the simulation.

For each extracted feature, a 3D ellipsoid glyph approximates the spatial distribution of the merged subvolumes by encoding average position, volume alignment, and size. The encoding is obtained by orienting ellipsoid axes according to the eigenvectors of the covariance matrix of the subvolumes' positions. The size of the ellipsoid is determined by the respective eigenvalues. Additionally, color can be used to communicate additional information (e.g., the kind of entities involved in the feature specification).

The ellipsoids are placed in a three dimensional representation of the simulation space. We use 3-band contours to render the ellipsoid glyphs. Such "open" glyphs help reducing occlusion of features. Figure 2 shows an example with several features. Red and blue ellipsoids stand for regions with high concentration of two different proteins of a bi-stable biochemical system (see [7] 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 structural relationships between features of consecutive time steps, which would otherwise remain hidden in the data. 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 [26].

In order to track features, one has to solve the correspondence problem. Two approaches to this problem exist: attribute correspondence and region correspondence [4]. Attribute correspondence is faster to compute, but less precise. Vice versa, region correspondence takes a bit longer to calculate, but yields more precise results.

4

(a) Ellipsoids at time step t_i (b) Ellipsoids for $k = 2$ time steps   (c) Feature graph for $n = 10$ time steps   (d) Ellipsoids and graph combined
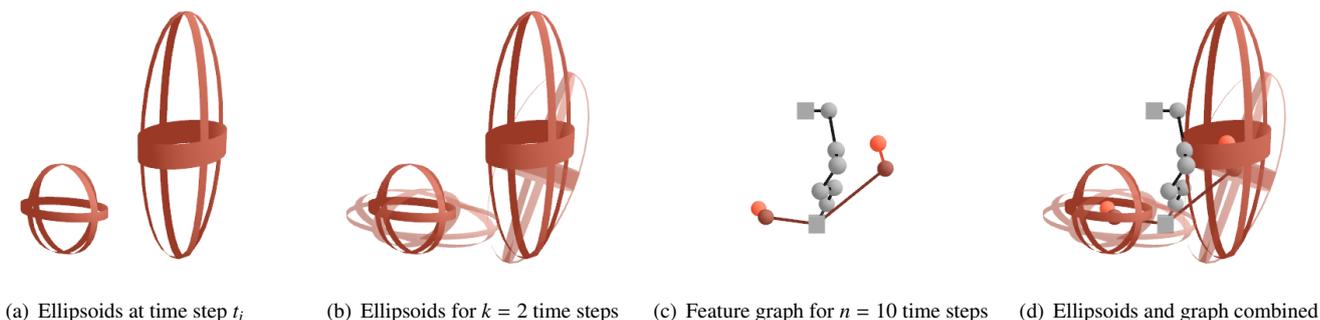
Figure 4: Visualization of two merging features. (a) + (b) Spatial convergence of features can be observed when showing ellipsoids for a small number of k time steps. (c) The point where features merge and the upward movement of the resulting feature can be investigated using the feature graph of n time steps ($n > k$). (d) The combination of both representations provides detailed spatial information of the temporal development in a narrow time frame (k), while the graph adds structural and temporal context to the view for a larger time frame (n).

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., [27]) to track features and detect events.

The result of the tracking algorithm can be interpreted as a *feature graph* (or event graph, see [4, 28, 29]) in which nodes represent features and edges connect related features. Paths in the graph link 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 lay it out in a time-line-like fashion. Along the horizontal axis we show the individual time steps. Each time step is associated with a layer that contains all features extracted from that time step properly stacked along the vertical axis (see Figure 3). The Sugiyama layout algorithm [30] is applied to minimize edge crossings. The visualization uses node size to encode feature "size" (e.g., volume, concentration). Connected nodes in the graph can again be colored depending on additional information (e.g., dominant type of protein). While the graph structure already indicates events (i.e., split or merge), we emphasize them via small symbols: dot and cross for birth and death, oriented triangles for split and merge, and star for simultaneous split and merge.

As illustrated in Figure 3, one can easily spot bigger features, and by following paths of 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. On the other hand, showing all the features of all time steps as 3D ellipsoid glyphs in space is likely to overcrowd the display. Therefore, we propose a novel composite visualization that brings together ellipsoid glyphs and the feature graph in the 3D simulation space.

We suggest showing ellipsoid glyphs for a limited number of successive time steps $[t_i, t_{i+k}]$. To avoid visual clutter k should be kept small (e.g., 2-4). Further we embed a 3D representation of the feature graph into the simulation space. Again to reduce clutter, we restrict the graph to a range of time steps $[t_i, t_{i+n}]$. As the graph representation takes less space than the ellipsoid glyphs n can be larger than k. As illustrated in Figure 4 the combination of 3D ellipsoid glyphs and 3D feature graph conveys spatial, temporal, and structural aspects in a single view. An advantage of this design is that ellipsoid glyphs provide the details in the time interval of $[t_i, t_{i+k}]$, while the graph provides a preview of where features will be located in $[t_{i+k}, t_{i+n}]$. Adjusting i, k, and n enables the user to navigate through time and to control the visual density of the visualization.

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 additional interaction mechanisms.

3.4. Interaction

In addition to the interactive means for the feature specification, our approach incorporates interaction tools to support the exploration and analysis process. We consider classic brushing and linking, temporal filtering as well as dedicated selection mechanisms that exploit the structural information inherent in the feature graph.

Our solution communicates spatial, temporal, and structural aspects of the same data in the 3D view of the simulation space and the abstract 2D view of the tracking graph. Both views are linked to assist users in mentally connecting the different data aspects.

Temporal filtering is carried out with the help of the gray frame depicted in Figure 3. The left and right side of the frame serve as interaction handles that can be used to narrow or expand the time range to be considered for the spatial visualiza-
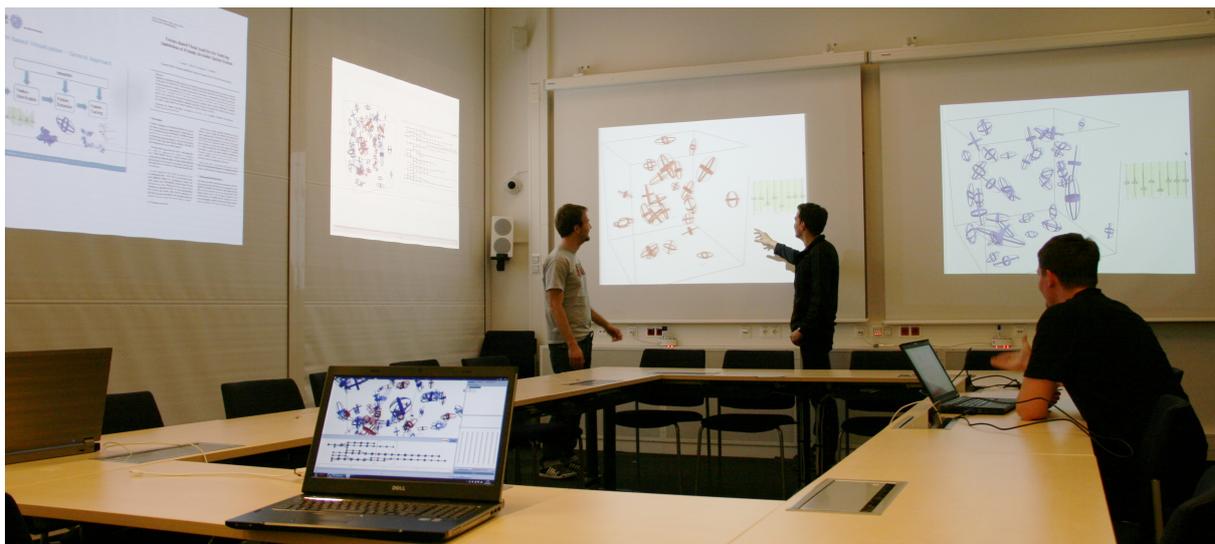
Figure 5: Feature-based visual analytics of biochemical simulation data in a smart meeting room. Multiple displays show multiple views from different sources. Documents provide additional context information. Two users discuss the data visualized on a large displays, a third user compares the data with his private views.

tion. Features beyond the selected time range (i.e., past or future time steps) can be dimmed or omitted altogether.

On top of temporal filtering and 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. These selections can be further restricted to capture only future or past features. The different selection modes can be combined to form any subset of features with just a few clicks.

These interaction facilities round off our feature-based visual analytics approach. Now that the conceptual part of our work has been explained, we will next shed some light on an implementation of the approach in a multi-view multi-display visualization environment.

4. Multi-View Multi-Display Implementation

A standard desktop implementation has been developed to demonstrate the potential of the feature-based approach. While the simulation experts acknowledged its usefulness, they were also concerned about its application in a real scenario. What the experts were longing for was an implementation that would allow them to compare multiple simulations of different models carried out using different parametrizations. As a single desktop application cannot serve this purpose, we are developing an implementation for a smart meeting room.

Smart meeting rooms are intelligent environments that aim to assist people in their various tasks [31]. In terms of visual analytics applications, smart meeting rooms have the advantage that they offer multiple displays that can show multiple views from different sources in a coordinated fashion. Large public and regular private displays support collaborative working and discussion scenarios. With these characteristics, a smart meeting room is an excellent environment for feature-based visual analytics of multi-faceted biochemical simulation data.

Our implementation builds upon previous work by Radloff et al. [32]. A view grabber captures visual representations from different applications, and so from (possibly multiple instances of) our software. These views are enriched with meta data that assist in the later layout of the views in the environment. The layout is handled by a component called smart view management. Based on the current situation (e.g., location of users, available displays, number of views), the smart view management suggests a suitable layout of the views that need to be displayed. Users can interactively refine the layout if special arrangements are necessary.

A big advantage of this approach is that additional views can be generated and added to the ensemble at any time. So if a different visualization is needed to compare hypothesis against alternative perspectives, a new instance of our feature-based visual analytics solution (or any other compatible visualization tool) can be started to generate the needed visual representations. These are then automatically integrated into the present layout. Moreover, complementary information, such as documentations or scientific publications and even additional tools, such as model editors or simulators can be coupled in the visual analytics scenario.

Figure 5 depicts how a discussion of biochemical simulation data might look like in a smart meeting room. Two central projections show features related to two different proteins of the simulation. Two auxiliary projections to the left show more details about the simulation including the tracking graph and a scientific publication. In the foreground one can see a notebook that serves as the source for a part of the visual information shown in the room's projections.

With this implementation we provide the simulation experts with a platform that not only helps them to gain insight into a single data set, but also to bring together and understand data from multiple sources and to discuss their hypotheses with other experts.
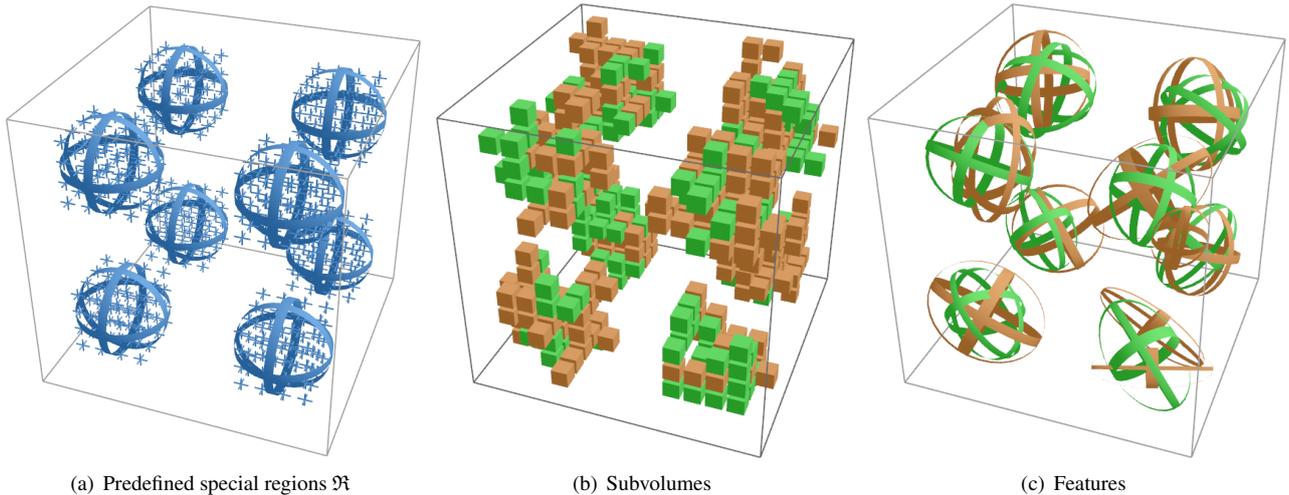
| (a) Predefined special regions $\mathfrak{R}$ | (b) Subvolumes | (c) Features |

Figure 6: (a) Predefined special regions $\mathfrak{R}$. (b) Showing subvolumes with high amounts of proteins B (brown) and C (green) indicates correspondence with $\mathfrak{R}$, but visual confirmation is difficult due to occlusion. (c) The feature-based visualization makes it easier to confirm that proteins B and C are concentrated at $\mathfrak{R}$.

5. Use case

Next we will describe a concrete example of how simulation experts used our approach to test a model simulation with two alternative configurations.
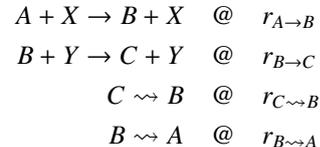
5.1. Biological Background

Biological cells respond to changes in their environment via multi-step processes consisting several reactions. For example, activation or deactivation of proteins may involve proteins binding or unbinding a phosphate group, or two proteins binding to each other. Heinrich et al. [33] describe the common pattern of activation of a receptor in the cell membrane, which activates an intracellular protein, which may activate others, which ultimately trigger, for example, gene transcription inside the nucleus and thus production of even other proteins.

Here, the focus is on an aspect of these signaling cascades where the spatial distribution of proteins may be relevant. For example, it has been observed for receptor proteins on a membrane that their rate of binding to each other depends on certain regions with different properties, so-called lipid rafts [34]. As another example, for bone cells growing on structured titanium surfaces, differences in protein presence could be observed above regions of the cell that touch the surface and those that do not [35]. Starting from the two examples of receptor dimerization in lipid rafts and integrin receptor complex formation in bone cells, the aim is to investigate more generally (and in a three-dimensional setting) how the spatial distribution of one mediating protein in a two-step process can influence the distribution of the final product when there is a homogeneous distribution of the initial substrate.

5.2. Model, Hypotheses, and Visual Analysis

The biological process studied here is a two-step modification process according to the following rules. A protein, whose initial form is called A, is modified by protein X to an intermediate form B. This reaction occurs at a certain rate, called $r_{A \to B}$.

In another reaction, the intermediate B is modified by protein Y to the form called C at rate $r_{B \to C}$. In both biological scenarios outlined in Section 5.1 it is a reasonable assumption that the second reaction happens much slower than the first one (i.e. $r_{A \to B} \gg r_{B \to C}$). Both modifications are reversible, where the reverse reactions do not require any additional protein or other model entity. In summary, the verbalized rules can be captured more formally as:

$$
\begin{aligned}
A + X &\to B + X & @ & \quad r_{A \to B} \\
B + Y &\to C + Y & @ & \quad r_{B \to C} \\
C &\rightsquigarrow B & @ & \quad r_{C \rightsquigarrow B} \\
B &\rightsquigarrow A & @ & \quad r_{B \rightsquigarrow A}
\end{aligned}
$$

A spatial simulation in ML-Space [8] is carried out to study this model. Starting with the participating proteins being distributed homogeneously across space, the simulation experts are interested in how the simulation can give rise to a situation where A is still roughly homogeneously distributed, but C is concentrated in certain known regions $\mathfrak{R}$. For simplicity, $\mathfrak{R}$ are considered to be 2^3 or 3^3 smaller cubic regions placed on a regular cubic grid as shown in Figure 6(a).

The visual analytics goal is to confirm (or reject) visually that specific simulation configurations lead to the desired simulation outcome. The simulation experts hypothesize that two alternative alterations of the base configuration may be used:

1. **Slow down B and C:** Simulator rules are deployed to slow down the diffusion of substrate and product proteins B and C in $\mathfrak{R}$. Further, either the reaction A to B or the reaction B to C is made significantly slower than the respective other. The slowed reaction takes place predominantly in areas where the reaction's substrate stays around longer before diffusing away.
2. **Confine X:** Simulator rules are deployed to confine protein X to the regions $\mathfrak{R}$. That is X may appear in and around $\mathfrak{R}$, but is not allowed to diffuse into areas outside of $\mathfrak{R}$.
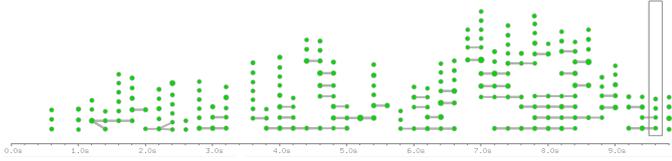
Figure 7: Tracking graph of features related to high concentration of C. As features are only short-lived, one can conclude that simulations using the *slow-down* configuration yield no stable accumulations of C.

(a) $r_{B\rightsquigarrow A}$ < $r_{C\rightsquigarrow B}$ (b) $r_{B\rightsquigarrow A}$ > $r_{C\rightsquigarrow B}$

Figure 8: Visual comparison of feature size at different rates of reversion. Features related to molecules of B and C are shown in brown and green, respectively.

Simulations are run using the hypothesized configurations and the resulting data are analyzed with the help of our feature-based visual analytics solution.

First we examine simulations based on *slowdown*. With little or no slowdown, C does not localize to the special regions $\mathfrak{R}$, but occurs randomly distributed. This can already be guessed from looking at the particle counts alone (even without feature extraction). But it is even more obvious when looking at the tracking graph of features referring to high concentrations of C as shown in Figure 5.2. There are many small features that exist only very shortly. No longer-lived features appear in the feature graph. Even in simulation runs with a higher slowdown ($\approx \frac{1}{4}$ of the original diffusion), no spatially stable features of C would surface. This means that the expected situation in the distribution of the final product does not emerge from simulations based on the configuration with *slowdown*.

Let us now take a look at simulations based on *confinement* of X to the special regions $\mathfrak{R}$. As the feature-based visualization reveals, confining X results in the desired simulation outcome. Figures 6(b) and 6(c) show that features related to high concentrations of proteins B (brown) and C (green) appear in areas corresponding to the predefined special regions $\mathfrak{R}$ indicated in Figure 6(a).

With the *confinement* configuration confirmed, the simulation experts studied the model more closely for further insight. One objective was to investigate the influence of the reaction rates. The feature-based visualization in Figure 8(a) shows that features related to B and C are roughly of equal size when the rate $r_{B\rightsquigarrow A}$ at which B reverts to A is lower than the rate $r_{C\rightsquigarrow B}$ at which C reverts to B. However, when B reverts to A faster than C reverts to B, features of C can be larger than those of B as can be seen in Figure 8(b). In biological terms, this means that a signal that triggers only a rather transient and/or local effect (A to B) can not only be sustained longer, but also be transmitted farther if there is a second, more durable step (B to C) enabled by the former.

Summing up, visualizing the output data of simulations based on two different configurations allowed the simulation experts to discard one of the two as the source of expected biological behavior. The visualization further confirmed that when one component of a two-step process is spatially inhomogeneously distributed, the result of the process is so as well. While this is not surprising, the results also indicate that the region where the product is located can be larger or smaller than that of the spatially constrained other components in the process, depending on the involved reversion rates.
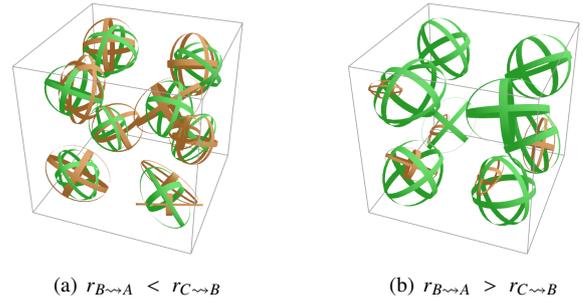
Several purposes of signaling cascades were known before. For instance, Heinrich et al. [33] report on amplification w.r.t. magnitude or sustaining a signal over time. With the feature-based visual analytics approach, it could be confirmed that they can also distribute signals spatially.

6. Preliminary User Feedback and Discussion

Our feature-based approach addresses concrete requirements of simulation experts. During the development of the concept and the implementation of the software, we constantly collected informal feedback from our users. In this section, we recap their impressions and opinions briefly and discuss advantages and potential problems of the current state of our solution.

Since raw data analysis had been the standard way of working with simulation data, it is not surprising that the general approach of high-level feature visualization was overall well received and got quite positive feedback.

The simulation experts welcomed the flexible specification of user-defined features, as it allows them to investigate different kinds of simulated processes with varying simulation settings. The integration of the Laplace operator to facilitate the specification of features representing high or low particle concentrations has proven to be useful. However, finding suitable thresholds of interesting particle concentrations and deciding on connectedness criteria for region merging requires an understanding of the underlying feature extraction process. Hence, new users need a short introduction and training phase. Interestingly, once the users had mastered the analytic part of the feature-based pipeline, there were hardly any problems in interpreting the resulting visualization itself.

Especially the ellipsoidal abstraction of the features was positively critiqued. However, we had to warn our users that, as a kind of simplified representation of spatial properties, ellipsoids cannot handle all simulation scenarios equally well. In fact, the calculations used to construct the ellipsoids are a good compromise to describe the overall volume distribution of particles. However, when a feature is to be interpreted as a surface of a region, ellipsoids are not well-suited. For example, extracted ellipsoids might intersect, while the exact region surfaces do not. This can be the case when features form network-like structures, or when large, separated features are close to

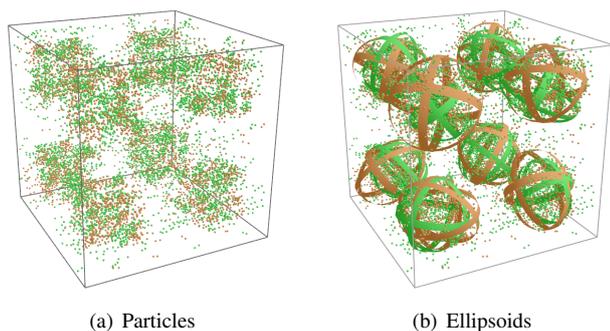<div style="text-align:center">(a) Particles   (b) Ellipsoids</div>

Figure 9: Comparison of raw data and ellipsoids. (a) Illustration of the raw data as individual particles. (b) Ellipsoids represent compact features quite well.

each other. The occurrence of such cases highly depends on the simulated system and the specified features. If a majority of subvolumes is classified as interesting, it is very likely that the extracted ellipsoids cover the entire simulation space and hence are not very useful. If, on the other hand, the focus is on high or low particle concentrations, the corresponding features are typically compact and spatially separated. In such cases, ellipsoids work out quite nicely as illustrated in Figure 9.

The integration of our approach in a smart meeting room had mixed receptions. In the first place, great potential was acknowledged. Multiple large displays enable users to analyze several views at the same time. This was mentioned to be particularly helpful when using several spatial views in order to directly compare differently parametrized simulation runs, alternative feature definitions, or customized visualization settings. In contrast to the desktop solution, the smart meeting room facilitates displaying multiple views without scaling them in size or switching between applications, which helps in recognizing details. On the downside, however, usability issues surfaced in the initial tests, especially when a lot of views were used. Most issues concern the placement and grouping of views on the available displays. While manually arranging views on the displays is a tedious task, the employed automatic layout mechanisms generated only suboptimal results on several occasions. For example, when a user wants to tune a feature definition, it would be helpful to automatically arrange the spatial view next to the histograms. In contrast, when extracted features of different simulations need to be compared, histograms are less important and views visualizing the spatial data should be shown together. While the current implementation allows a user to configure layout strategies and grouping of views manually, it would be much more intuitive if such constraints could be extracted automatically. Dealing with these kinds of questions will be a task for future work.

7. Conclusion

In this work, we presented visual analytics support for the investigation of simulations of biochemical systems. The analytic component of our solution is largely inspired by the classic feature-based strategy, 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 detail. 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 implemented in a multi-view multi-display smart meeting room. Simulation experts have used our approach to test different hypotheses about the simulation of biochemical systems. 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 [2] or protein trajectories [10]). 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. Under the assumption that the usability issues discussed in the previous section can be tackled, we believe that an integrated visual analytics environment with low-level data visualization and high-level features will enable new ways of gaining insight into simulations of biochemical processes.

References

[1] Doleisch H. SIMVIS: Interactive Visual Analysis of Large and Time-dependent 3D Simulation Data. In: Proceedings of the Winter Simulation Conference (WSC). IEEE Press; 2007, p. 712–20. doi:10.1145/1351542.1351674.

[2] Unger A, Schumann H. Visual Support for the Understanding of Simulation Processes. In: Proceedings of the Pacific Visualization Symposium (PacificVis). IEEE Computer Society; 2009, p. 57–64. doi:10.1109/PACIFICVIS.2009.4906838.

[3] Tufte ER. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press; 1983.

[4] Reinders F, Post FH, Spoelder HJ. Visualization of Time-Dependent Data with Feature Tracking and Event Detection. The Visual Computer 2001;17(1):55–71. doi:10.1007/PL00013399.

[5] Eichner C, Bittig A, Schumann H, Tominski C. Feature-Based Visual Analytics for Studying Simulations of Dynamic Bi-Stable Spatial Systems. In: Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA). Eurographics Association; 2013, p. 25–9. doi:10.2312/PE.EuroVAST.EuroVA13.025-029.

[6] Stundzia AB, Lumsden CJ. Stochastic Simulation of Coupled Reaction-Diffusion Processes. Journal of Computational Physics 1996;127(1):196–207. doi:10.1006/jcph.1996.0168.

[7] Elf J, Ehrenberg M. Spontaneous Separation of Bi-Stable Biochemical Systems into Spatial Domains of Opposite Phases. Systems Biology 2004;1:230–6. doi:10.1049/sb:20045021.

[8] Bittig AT, Haack F, Maus C, Uhrmacher AM. 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). ACM; 2011, p. 161–70. doi:10.1145/2037509.2037533.

[9] Gribble CP, Ize T, Kensler A, Wald I, Parker SG. A Coherent Grid Traversal Approach to Visualizing Particle-Based Simulation Data. IEEE Transactions on Visualization and Computer Graphics 2007;13(4):758–68. doi:10.1109/TVCG.2007.1059.

[10] Luboschik M, Tominski C, Bittig AT, Uhrmacher AM, Schumann H. Towards Interactive Visual Analysis of Microscopic-Level Simulation Data. In: Proceedings of the Annual SIGRAD Conference, Special Theme: InteractiveVisual Analysis of Data. Linkping University Electronic Press; 2012, p. 91–4.

[11] Burger K, Kondratieva P, Kruger J, Westermann R. Importance-Driven Particle Techniques for Flow Visualization. In: Proceedings of the Pacific Visualization Symposium (PacificVis). IEEE Computer Society; 2008, p. 71–8. doi:10.1109/PACIFICVIS.2008.4475461.

[12] Krüger J, Kipfer P, Kondratieva P, Westermann R. A Particle System for Interactive Visualization of 3D Flows. IEEE Transactions on Visualization and Computer Graphics 2005;11(6):744–56. doi:10.1109/TVCG.2005.87.

[13] Kehrer J, Hauser H. Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey. IEEE Transactions on Visualization and Computer Graphics 2013;19(3):495–513. doi:10.1109/TVCG.2012.110.

[14] Post FH, Vrolijk B, Hauser H, Laramee RS, Doleisch H. The State of the Art in Flow Visualisation: Feature Extraction and Tracking. Computer Graphics Forum 2003;22(4):775–92. doi:10.1111/j.1467-8659.2003.00723.x.

[15] Tzeng FY, Ma KL. Intelligent Feature Extraction and Tracking for Visualizing Large-Scale 4D Flow Simulations. In: Proceedings of the ACM/IEEE Conference on Supercomputing (SC). IEEE Computer Society; 2005,doi:10.1109/SC.2005.37.

[16] Anand A, Dang TN, Wilkinson L. Visual Pattern Discovery Using Random Projections. In: Proceedings of the Conference on Visual Analytics Science and Technology (VAST). IEEE Computer Society; 2012, p. 43–52. doi:10.1109/VAST.2012.6400490.

[17] 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). IEEE Computer Society; 2012, p. 73–82. doi:10.1109/VAST.2012.6400487.

[18] Jänicke H, Böttinger M, Scheuermann G. Brushing of Attribute Clouds for the Visualization of Multivariate Data. IEEE Transactions on Visualization and Computer Graphics 2008;14(6):1459–66. doi:10.1109/TVCG.2008.116.

[19] Wong PC, Foote H, Adams D, Cowley W, Thomas J. Dynamic Visualization of Transient Data Streams. In: Proceedings of the Symposium on Information Visualization (InfoVis). IEEE Computer Society; 2003, p. 97–104. doi:10.1109/INFVIS.2003.1249014.

[20] Rohrdantz C, Hao MC, Dayal U, Haug LE, Keim DA. Feature-Based Visual Sentiment Analysis of Text Document Streams. ACM Transactions on Intelligent Systems and Technology 2012;3(2):26:1–26:25. doi:10.1145/2089094.2089102.

[21] Höferlin B, Höferlin M, Heidemann G, Weiskopf D. Scalable Video Visual Analytics. Information Visualization 2013;doi:10.1177/1473871613488571; to appear.

[22] Grottel S, Reina G, Vrabec J, Ertl T. Visual Verification and Analysis of Cluster Detection for Molecular Dynamics. IEEE Transactions on Visualization and Computer Graphics 2007;13(6):1624–31. doi:10.1109/TVCG.2007.70614.

[23] 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). Eurographics Association; 2003, p. 239–48.

[24] 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: Springer-Verlag; 2011, p. 619–28. doi:10.1007/978-3-642-24028-7_57.

[25] Walsum T, Post FH, Silver D, Post FJ. Feature Extraction and Iconic Visualization. IEEE Transactions on Visualization and Computer Graphics 1996;2(2):111–9. doi:10.1109/2945.506223.

[26] Samtaney R, Silver D, Zabusky N, Cao J. Visualizing Features and Tracking Their Evolution. Computer 1994;27(7):20–7. doi:10.1109/2.299407.

[27] Silver D, Wang X. Tracking and Visualizing Turbulent 3D Features. IEEE Transactions on Visualization and Computer Graphics 1997;3(2):129–41. doi:10.1109/2945.597796.

[28] 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). IEEE Computer Society; 2003, p. 103–10. doi:10.1109/PVGS.2003.1249048.

[29] Bremer PT, 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 2011;17(9):1307–24. doi:10.1109/TVCG.2010.253.

[30] Sugiyama K, Tagawa S, Toda M. Methods for Visual Understanding of Hierarchical System Structures. IEEE Transactions on Systems, Man and Cybernetics 1981;11(2):109–25. doi:10.1109/TSMC.1981.4308636.

[31] Yu Z, Nakamura Y. Smart Meeting Systems: A Survey of State-of-the-Art and Open Issues. ACM Computing Surveys 2010;42(2):8:1–8:20. doi:10.1145/1667062.1667065.

[32] Radloff A, Luboschik M, Schumann H. Smart Views in Smart Environments. In: Proceedings of the International Symposium on Smart Graphics (SG). Springer; 2011, p. 1–12. doi:10.1007/978-3-642-22571-0_1.

[33] Heinrich R, Neel BG, Rapoport TA. Mathematical Models of Protein Kinase Signal Transduction. Molecular Cell 2002;9(5):957–70.

[34] Haack F, Burrage K, Redmer R, Uhrmacher AM. Studying the Role of Lipid Rafts on Protein Receptor Bindings with Cellular Automata. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2013;doi:10.1109/TCBB.2013.40; to appear.

[35] Bittig AT, Uhrmacher AM, Matschegewski C, Nebe JB. Spatial Simulation of Actin Filament Dynamics on Structured Surfaces. In: Proceedings of the Winter Simulation Conference (WSC). IEEE; 2012,Poster.