Computers & Graphics 38 (2014) 18–26

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Visual Analytics

# 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

ARTICLE INFO

*Article history:*
Received 21 June 2013
Received in revised form
21 August 2013
Accepted 10 September 2013
Available online 21 September 2013

*Keywords:*
Visual analytics
Feature detection
Feature tracking
Feature visualization
Simulation

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.


## 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 simulation 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

* Corresponding author. Tel.: +49 3814987435.
*E-mail address:* christian.eichner@uni-rostock.de (C. Eichner).

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.09.001

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 a 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.

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 data sets 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.