Visual Analysis of High-Dimensional Point Clouds using Topological Abstraction

Dissertation, 2016

Patrick Oesterling

This thesis is about visualizing a kind of data that is trivial to process by computers but difficult to imagine by humans because nature does not allow for intuition with this type of information: high-dimensional data. Such data often result from representing observations of objects under various aspects or with different properties. In many applications, a typical, laborious task is to find related objects or to group those that are similar to each other. One classic solution for this task is to imagine the data as vectors in a Euclidean space with object variables as dimensions. Utilizing Euclidean distance as a measure of similarity, objects with similar properties and values accumulate to groups, so-called clusters, that are exposed by cluster analysis on the high-dimensional point cloud. Because similar vectors can be thought of as objects that are alike in terms of their attributes, the point cloud's structure and individual cluster properties, like their size or compactness, summarize data categories and their relative importance.

The contribution of this thesis is a novel analysis approach for visual exploration of high-dimensional point clouds without suffering from structural occlusion. The work is based on implementing two key concepts:

The first idea is to discard those geometric properties that cannot be preserved and, thus, lead to the typical artifacts. Topological concepts are used instead to shift away the focus from a point-centered view on the data to a more structure-centered perspective. The advantage is that topology-driven clustering information can be extracted in the data's original domain and be preserved without loss in low dimensions.

The second idea is to split the analysis into a topology-based global overview and a subsequent geometric local refinement. The occlusion-free overview enables the analyst to identify features and to link them to other visualizations that permit analysis of those properties not captured by the topological abstraction, e.g. cluster shape or value distributions in particular dimensions or subspaces. The advantage of separating structure from data point analysis is that restricting local analysis only to data subsets significantly reduces artifacts and the visual complexity of standard techniques. That is, the additional topological layer enables the analyst to identify structure that was hidden before and to focus on particular features by suppressing irrelevant points during local feature analysis.

This thesis addresses the topology-based visual analysis of high-dimensional point clouds for both the time-invariant and the time-varying case. Time-invariant means that the points do not change in their number or positions. That is, the analyst explores the clustering of a fixed and constant set of points. The extension to the time-varying case implies the analysis of a varying clustering, where clusters appear as new, merge or split, or vanish. Especially for high-dimensional data, both tracking – which means to relate features over time – but also visualizing changing structure are difficult problems to solve.