Visualizing High-Dimensional Data: Advances in the Past Decade

Shusen Liu, Dan Maljovec, Bei Wang, Peer-Timo Bremer, and Valerio Pascucci, *Member, IEEE*

Abstract—Massive simulations and arrays of sensing devices, in combination with increasing computing resources, have generated large, complex, high-dimensional datasets used to study phenomena across numerous fields of study. Visualization plays an important role in exploring such datasets. We provide a comprehensive survey of advances in high-dimensional data visualization that focuses on the past decade. We aim at providing guidance for data practitioners to navigate through a modular view of the recent advances, inspiring the creation of new visualizations along the enriched visualization pipeline, and identifying future opportunities for visualization research.

Index Terms—Taxonomy, high-dimensional data, multidimensional data, visualization, data models, computational modeling

1 Introduction

With the ever-increasing amount of available computing resources and sensing devices, our ability to collect and generate a wide variety of large, complex datasets continues to grow. High-dimensional datasets show up in numerous fields of study, such as economics, biology, chemistry, political science, astronomy, and physics, to name a few. Their wide availability, increasing size, and complexity have led to new challenges and opportunities for their effective visualization. For example, genomic microarrays in biology [1], [2], spectrometry data in air quality research [3], simulation parameters in nuclear safety engineering [4], and chemical compositions in combustion simulations [5] can all be mapped to high-dimensional spaces (with a few dozen to several hundreds of dimensions) for exploration.

On the other hand, the physical limitations of display devices and our visual systems prevent the direct display and rapid recognition of structures with dimensions higher than two or three. In the past decade, a variety of approaches have been introduced to visually convey high-dimensional structural information by utilizing low-dimensional projections or abstractions: from dimension reduction to visual encoding, and from quantitative analysis to interactive exploration. A number of surveys have focused on different aspects of high-dimensional data visualization, such as parallel coordinates [6], [7], quality measures [8], clutter reduction [9], visual data mining [10], [11], [12], and interactive techniques [13]. Multivariate scientific datasets have also been investigated in [14], [15], while other surveys [16], [17], [18] have focused on the various aspects of visual encoding techniques. These papers provide a valuable summary of existing techniques and inspiring discussions of future directions in their respective domains. However, few surveys in the past decade have aimed at providing a general, coherent, and unified picture that addresses the full spectrum of techniques for visualizing high-dimensional data.

In this work, we strive to provide a broad survey of advances in high-dimensional data visualization over the past decade (even though the focus is on the last decade, the search extends to more than 15 years), with the following objectives: providing guidance for data practitioners to navigate through a modular view of the recent advances, allowing the creation of new visualizations along the enriched visualization pipeline, and identifying opportunities for future visualization research.

A high-dimensional dataset can be described through the perspective of the *range* and *domain* of a function, which provides a unified view of several related but different types of datasets. In this survey, a dataset with more than three domain or range attributes is considered high-dimensional.

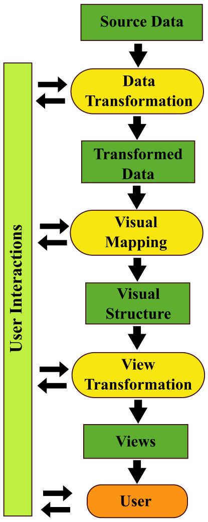
Our contributions are as follows. We propose a categorization of recent advances based on the visualization pipeline [19], enriched with customized classifications (Fig. 1, Section 2) to highlight the common operations in each stage of the pipeline (Sections 3, 4, and 5). We further assess the interplay between user interaction and the visualizaiton pipeline and summarize the prominent interaction patterns in this context (Fig. 7, Section 6). Finally, we provide a discussion of emerging research directions in connection with high-dimensional data visualization (Section 7), as well as a summary and reflection on our categorization (Section 8). This paper includes and extends our earlier survey [20] by enriching existing topics, deliberating about emerging ones and reflecting on the surveying process.

- *S. Liu, D. Maljovec, B. Wang, and V. Pascucci are with the Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT 84112. E-mail: {shusenl, maljovec, beiwang, pascucci}@sci.utah.edu.*
- *P.-T. Bremer is with Lawrence Livermore National Laboratory, Livermore, CA 94550. E-mail: bremer5@llnl.gov.*

Manuscript received 1 Apr. 2016; revised 9 Nov. 2016; accepted 9 Dec. 2016. Date of publication 15 Dec. 2016; date of current version 1 Feb. 2017. Recommended for acceptance by E. Gobbetti.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2016.2640960

2 Survey Method and Categorization

We have conducted a thorough literature review based on relevant works from major visualization venues, namely Visweek, EuroVis, PacificVis, and the journal *IEEE Transactions*



Data Transformation	Dimension Reduction	Subspace Clustering	Regression Analysis	Topological Data Analysis
	linear projection [21], [22], non-linear DR [23], [24], Control Points Projection [25], [26], Distance Metric [27], [28], Precision Measures [29], [30]	Dimension Space Exploration [2], [31], [32], Subset of Dimension [33], [34], Non-Axis-Parallel Subspace [35], [36], [37]	Optimization Design Steering [38], [39], [40], Structural Summaries [5], [41]	Morse-Smale Complex [42], [43], [44], [45] Reeb Graph [46], [47], [48] Contour Tree [49], [50], Topological Features [51], [52]

Visual Mapping	Axis Based	Glyphs	Pixel-Oriented	Hierarchy Based	Animation	Evaluation
	Scatterplot Matrix [53], Parallel Coordinate [54], Radial Layout [55], Hybrid Construction [56], [57], [58], [59]	Per-Element Glyphs [60], [61] [62], [63], Multi-Object Glyphs [64], [65], [66]	Jigsaw Map [67], Pixel Bar Charts [68], Circle Segment [69], Value & Relation Display [70]	Dimension Hierarchy [71], Topology-based Hierarchy [72], [73], Others [74], [75]	GGobi [76], TripAdvisorND [77], Rolling the Dice [78]	Scatterplot Guideline [79], [80], PCPs Effectiveness [81], Radical Layout [82], Animation [83]

View Transformation	Illustrative Rendering	Continuous Visual Representation	Accurate Color Blending	Image Space Metrics
	Illustrative PCP [84], Illuminated 3D Scatterplot [85], PCP Transfer Function [86] Magic Lens [87], [88]	Continuous Scatterplot [89], [90], Continuous Parallel Coordiante [91], [92], Splatterplots [93] Splatting for PCPs [94]	Hue-Preserving Blending [95], Weaving vs. Blending [96]	Clutter Reduction [97], [98], Pargnostics [99], Pixnostic [100]

Fig. 1. Research categorization based on different stages of the visualization pipeline, with subcategories that reflect common approaches.

on Visualization and Computer Graphics (TVCG) from the period between 2000 and 2015. To ensure the survey covers the state-of-the-art, we further selectively searched through references within the initial set of papers. Beyond the visualization field, we also dedicated special attention to the exploratory data analysis techniques in the statistics community. Through such a rigorous search process, we have identified more than 200 papers that focus on a wide spectrum of techniques for high-dimensional data visualization. To help organize the large quantity of papers, we have produced an interactive survey website[1] that allows readers to select and filter papers through various tags. Due to the space limitation, not all works in the complete list (available through the survey website) are discussed in this survey.

As illustrated in Fig. 1, we base our main categorization on the three transformation stages of the visualization pipeline [19] (and its minor variation in [8]), namely, data transformation, visual mapping, and view transformation. Each category is enriched with customized subcategories that reflect common approaches. Instead of focusing on a complete coverage of relevant research, we strive to provide a broad overview of advances pertinent to high-dimensional data visualization while highlighting representative works, through the carefully designed subcategories, which can act as guidelines for interested reader to dive into more specific topics or techniques.

Data transformation (Section 3) corresponds to the analysis-centric methods such as dimension reduction, regression, subspace clustering, feature extraction, data sampling, and abstraction. *Visual mapping* (Section 4) emphasizes visual encoding tasks that transform the information from the data transformation stage for visual representation. This category includes visual encodings based on axes (e.g., scatterplots and parallel coordinate plots), glyphs, pixels, and hierarchical representations, together with animation and perception. *View transformation* (Section 5) methods focus on screen space and rendering. Examples from this stage include illustrative rendering for various visual structures, as well as screen space measures for reducing clutter or artifacts and highlighting important features.

This design allows us to easily classify the core contributions of vastly different methods that operate on entirely different objects, but at the same time, reveal their interconnections through the linked pipeline. Also, the pipeline-based categorization provides the reader with a modular view of the recent advances, allowing new systems to be configured based on possible options provided by the reviewed methods.

Interactivity is an integral part within each stage of the pipeline (Section 6), as illustrated in Fig. 1. Based on the amount of user interaction, we classify high-dimensional data visualization methods into three categories: computation-centric, interactive exploration, and model manipulation. The distinction between the latter two categories is made to emphasize a particular manipulation paradigm, where the underlying data model is altered based on interaction to reflect user intention.

Next, we identify two emerging fields of interest in Section 7. We survey related works in these areas in a context independent from the visualization pipeline in order to consolidate and highlight future directions of exploration. Finally, Section 8 serves to distill the key points of our survey.

3 Data Transformation

This section discusses in-depth the typical analysis techniques during data transformation, namely, dimension reduction, subspace clustering and regression analysis, as well as the emerging topic of topological data analysis. We focus particularly on their usages in visualization.

3.1 Data Value Type

Data transformation starts with input data. The attribute value type (e.g., nominal versus numerical) can greatly impact the choice and design of the visualization. In many applications, the value of the attributes is nominal in nature. However, most commonly available high-dimensional data visualization techniques are designed to handle numerical values only. When utilizing these methods for nominal data, information overlapping and stacking of visual elements

1. www.sci.utah.edu/~shusenl/highDimSurvey/website/. The site is developed based on the SurVis [21] framework.

usually exist. One way to address this challenge is to map the nominal values to numerical values [22] (e.g., as implemented in the XmdvTool [23]). Through such a mapping, each axis is used more efficiently, and the spacing becomes more meaningful. In the Parallel Sets work [24], the authors introduce a new visual representation that adapts the notion of parallel coordinates but replaces the data points with a frequency-based visual representation that is designed for nominal data. The Conjunctive Visual Form [25] allows users to rapidly query nominal values with certain conjunctive relationships through simple interactions. The Generalized Plot Matrix (GPLOM) [26] extends the Scatterplot Matrix (SPLOM) to handle nominal data. In recent work [27], Zhang et al. introduce the visual correlation analysis for both numerical and categorical data. In addition to the difference between nominal and numerical value type, data with a temporal dimension also requires different approaches. In most situations, the temporal dimension is analyzed separately, as demonstrated in TimeSpan [28].

3.2 Dimension Reduction

Dimension reduction is one of the fundamental techniques for analyzing and visualizing high-dimensional datasets. Dimension reduction techniques can be roughly divided into two major categories: linear dimension reduction and nonlinear dimension reduction (manifold learning).

Linear Projection. Linear projection uses linear transformation to project the data from high-dimensional to low-dimensional space. It includes many classical methods, such as Principal Component Analysis (PCA), Multidimensional Scaling (MDS), Linear Discriminant Analysis (LDA), and various factor analysis methods.

PCA [29] is designed to find an orthogonal linear transformation that maximizes the variance of the resulting embedding. PCA can be calculated by an eigen decomposition of the data's covariance matrix or a singular value decomposition of the data matrix. The interactive PCA (iPCA) [30] introduces a system that visualizes the results of PCA using multiple coordinated views. The system allows synchronized exploration and manipulations among the original data space, the eigenspace, and the projected space, which aids the user in understanding both the PCA process and the dataset. When visualizing labeled data, class separation is usually desired. Methods such as LDA aim to provide a linear projection that maximizes the class separation. The work by Koren et al. [31] generalizes PCA and LDA by providing a family of flexible linear projections to cope with different kinds of data.

Nonlinear Dimension Reduction. Nonlinear dimension reduction can occur in either a metric or nonmetric setting. The graph-based techniques are designed to handle *metric* inputs, such as Isomap [32], Locally Linear Embedding (LLE) [33], and Laplacian Eigenmap (LE) [34], where a neighborhood graph is used to capture local distance proximities and build a data-driven model of the space. The other group of techniques addresses nonmetric problems commonly referred to as nonmetric MDS or stress-based MDS by capturing nonmetric dissimilarities. The fundamental idea behind the nonmetric MDS is to minimize the mapping error directly through iterative optimizations. The well-known Shepard-Kruskal algorithm [35] begins by finding a monotonic transformation that maps the nonmetric dissimilarities to the metric distances, which preserves the rank-order of dissimilarity. Then, the resulting embedding is iteratively improved based on stress. The progressive and iterative nature of these methods has been exploited recently by Williams et al. [36], where the user is presented with a coarse approximation from partial data. The refinement is on-demand based on user inputs. Others rely on hybrid methods [37], [38] based upon stochastic sampling and interpolation to approximate the solution. t-SNE [39] has gained a lot of attention recently due to its effectiveness for visualizing high-dimensional data. It utilizes a probability distribution to encode the interpoint neighborhood information, and a mismatched distribution between high- and low-dimensional spaces is used to eliminate the unwanted attractive forces, therefore, resolving the crowding problem [39].

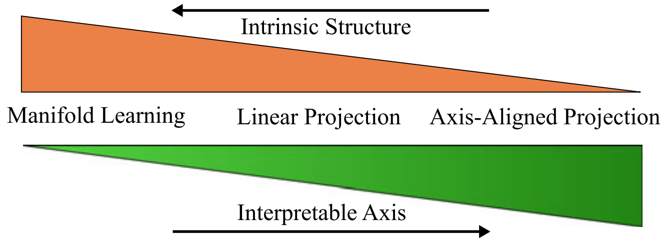


Fig. 2. A trade-off exists between the interpretablility of the axis and the intrinsic structure captured by the dimensionality reduction methods.

The trade-off among the different type of projections is illustrated in Fig. 2. The bivariate scatterplot (as in a scatterplot matrix) is most easily understood, since its axes directly correspond to the original dimensions. A linear projection [29], [31] generates interpretable embeddings (less so compared to a bivariate scatterplot), and the out-of-samples points can be easily projected to the same space. The nonlinear projection (manifold learning) approaches [32], [33], [34], on the other hand, allow the capture of more complex structures, but the resulting embedding can be extremely difficult to interpret.

Control Point Based Projection. For handling large and complex datasets, the traditional linear or nonlinear dimension reductions are limited by their computational efficiency. Some recent developments, e.g., [40], [41], [42], utilize a two-phase approach, where a set of control points (or anchor points) is projected first, followed by the projection of the rest of the points based on the location of the control points and preservation of local features. Such designs lead to a much more scalable system. Furthermore, the control points allow the user to easily manipulate and modify the outcome of the dimension reduction computation to achieve the desired results.

Distance Metric. For a given dimension reduction algorithm, a suitable distance metric is essential for the computation outcome as it is more likely to reveal important structural information. Brown et al. [43] introduce the distance function learning concept, where a new distance metric is calculated from the manipulation of point layouts by an expert user. In the *Explainers* work [44], the author attempts to associate a linear basis with a certain meaningful concept constructed based on user-defined examples. Machine learning techniques can then be employed to find a set of simple linear bases that achieve an accurate projection according to the prior examples. The structure-based analysis method [45] introduces a data-driven distance metric inspired by the perceptual processes of identifying distance relationships in parallel coordinates using polylines.

Dimension Reduction Precision Measure. One of the fundamental challenges in dimension reduction is assessing and measuring the quality of the resulting embeddings. Lee et al. introduce the ranking-based metric [46] that assesses the ranking discrepancy before and after applying dimension reduction. This technique is then generalized [47] and used for visualizing dimension reduction quality.

A projection precision measure is introduced in [48], where a local precision score is calculated for each point with a certain neighborhood size. In the distortion-guided exploration work [49], several distortion measures are proposed for different dimension reduction techniques, for which these measures aid in understanding the cause of highly distorted areas during interactive manipulation and exploration. For MDS, the stress can be used as a precision measure. Seifert et al. [50] further develop this idea by incorporating the analysis and visualization for better understanding of the localized stress phenomena. In recent work [51], Stahnke et al. introduce the notion of probing for examining the dimension reduction results. This approach not only reveals points with larger errors but also interactively considers locally correct representations of these points.

3.3 Subspace Clustering

Clustering is one of the most widely used data-driven analysis methods. Instead of providing an in-depth discussion of all clustering techniques, in this survey we focus on subspace clustering techniques that have a great impact on understanding and visualizing high-dimensional datasets. Compared to dimension reduction, which aims to compute one single embedding that best describes the structure of the data, subspace clustering helps identify multiple embeddings, each capturing a different aspect of the data, by clustering either the dimensions or the data points.

Dimension Space Exploration. Guided by the user, dimension space exploration methods interactively group relevant dimensions into subsets. The grouping allows us to better understand dimension relationships and to identify shared patterns among the dimensions. Turkay et al. introduce a *dual* visual analysis model [2] where both the dimension embedding and point embedding can be explored simultaneously. Their later improvement [52] allows for the grouping of a collection of dimensions as a *factor*, which permits effective exploration of the heterogeneous relationships among them. The Projection Matrix/Tree work [53] extends a similar concept to allow a recursive exploration of both the dimension space and data space. One recent advance [54] bridges the gap between the dimension space and the data space. By combining the dimension and element relationship and encoding them into a single matrix, the proposed approach produces a comprehensive map in which the data points are presented in the context of the variables. Several visual encoding methods also rely on the concept of dimension space exploration. These methods are discussed in Section 4.3.

Subsets of Dimensions. Compared to the dimension space exploration, where the user is responsible for identifying patterns and relationships, subspace clustering/finding methods automatically group related dimensions into clusters. Subspace clustering filters out the interferences introduced by irrelevant dimensions, allowing lower-dimensional structures to be discovered. These methods, such as ENCLUS [55], originate from the data mining and knowledge discovery community. They introduce some very interesting exploration strategies for high-dimensional datasets that can be particularly effective when the dimensions are not tightly coupled. The $TripAdvisor^{ND}$ [56] system employs a sightseeing metaphor for high-dimensional space navigation and exploration. It utilizes subspace clustering to identify the sights for the exploration. The subspace search and visualization work [57] utilizes the SURFING [58] algorithm to search the high-dimensional space and automatically identifies a large candidate set of interesting subspaces. In the work presented by Ferdosi et al. [59], morphological operators are applied on the density field generated from the (3D) PCA projection of the high-dimensional data for identifying subspace clusters.

Non-Axis-Aligned Subspaces. Instead of grouping the dimensions, which essentially creates axis-aligned linear subspaces, identifying non-axis-aligned linear subspaces is a more flexible alternative. Projection Pursuit [60] is one of the earliest works aimed at automatically identifying the interesting non-axis-aligned subspaces, where the projections are considered to be more interesting when they deviate more from a normal distribution. Recently, some advances have been made in the machine learning community to perform non-axis-aligned subspace clustering [61]. Instead of finding (possibly overlapping) clusters in axis-aligned subspaces defined by different dimensions combinations, the points are directly clustered together for sharing similar linear subspaces. In particular, this approach assumes the complex structure of the data can be approximated by a mixture of linear subspaces (of varying dimensions), and each of the linear subspaces corresponds to a set of points where their relationships can be approximately captured by the same linear subspace. Lehmann et al. [62] have recently introduced an interesting and different approach for identifying a set of distinct linear projections. By adopting a dissimilarity measure, they aim to remove duplicated data patterns by optimizing the dissimilarity among the selected projections. By utilizing random projection [63], Anand et al. [64] introduce an efficient subspace finding algorithm for data with thousands of dimensions. The algorithm generates a set of candidate subspaces through random projections and presents the top-scoring subspaces in an exploration tool.

3.4 Regression Analysis

Regression analysis for high-dimensional data is an extensive field of research on its own, and so, we focus only on the interplay between visualization and regression analysis.

Optimization and Design Steering. Pure optimization problems often are not the focus in the visualization community. What is more common are design steering methods for which, in addition to a multivariate input space, users have one or several output or response variables they want to explore (e.g., [65], [66]), where the results require a qualitative examination or are used to inform decisions.

HyperMoVal [67] is a software system used for validating regression models against actual data. It uses support vector regression (SVR) [68] to fit a model to high-dimensional data, highlights discrepancies between the data and the model, and computes sensitivity information on the model. The software allows for adding more model parameters to refine the regression to an acceptable level of accuracy. Berger et al. [65] utilize two types of regression models (SVR and nearest neighbor regression) to analyze a trade-off study in performance car engine design. Utilizing the predictive power of the regression, they are able to provide a guided navigation of the high-dimensional space centered around a user-selected focal point. The user adjusts the focal point through multiple linked views, and sensitivity and uncertainty information is encoded around the focal point.

Tuner [66] uses an automated adaptive sampling algorithm where a sparse sampling of the parameter space is refined by building a Gaussian Process Model (GPM) [69] and using adaptive sampling to focus additional samples in areas with either a high *goodness of fit* or high uncertainty. The software then relies heavily on user interaction to study the sensitivities with respect to each input parameter and steers the computation toward the user-defined optimal solution. Demir et al. [70] improve the effectiveness of GPMs by utilizing a block-wise matrix inversion scheme that can be implemented on the GPU, greatly increasing efficiency. In addition, their method involves progressive refinement of the GPM and can be halted at any point, if the improvement becomes insignificant.

Most of these methods convey sensitivity information through user exploration of the input space. In Section 4.2, explicit visual encodings for understanding sensitivity information are also discussed.

Structural Summaries. Researchers have also used regression to summarize data as in the works by Reddy et al. [71] and Gerber et al. [5]. Both approaches summarize the structures of the data via skeleton representations. Reddy et al. [71] use a clustering algorithm followed by construction of a minimum spanning tree of the cluster centroids in order to determine possible trends in the data. These trends are then fitted with principal curves [72] that go through the *medial-axis* of the data. HDViz [5], on the other hand, approximates a topological segmentation (for more details, see Section 3.5) and constructs an inverse linear regression for each segment of the data. In both examples, regression is used as a postprocessing step of the algorithms in order to present summaries of the extracted subsets of the data.

3.5 Topological Data Analysis

A crucial step in gaining insights from large, complex, high-dimensional data involves feature abstraction, extraction, and evaluation in the spatiotemporal domain for effective exploration and visualization. Topological data analysis (see [73], [74], [75], [76], [77], [78], [79] for seminal works and surveys), has provided efficient and reliable feature-driven analysis and visualization capabilities.

Topology in visualization covers many techniques dealing with multivariate data over a low-dimensional (e.g., 2, 3 or 4) spatiotemporal domain. This includes well-established research topics in vector and tensor field visualizations, and we defer discussion of such topics to the appropriate surveys [80], [81], [82], [83], [84], [85], [86]. Within this body of work, a few techniques have stated their applicability to arbitrary dimensional domain spaces [85], [87]; however very few applications exist for visualizing vector or tensor fields in high-dimensional spaces.

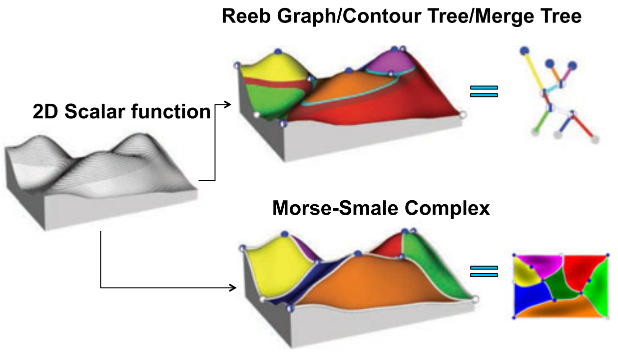


Fig. 3. Contour- and gradient-based topological structure of a 2D scalar function.

In our context, *topological data analysis* (TDA) is an emerging field of study that combines algebraic topology and other pure mathematical disciplines with computer science to describe the shape of data in a quantitative and mathematically rigorous fashion. Previous work in pure mathematics has focused on the study of topological spaces under smooth and continuous settings without computational considerations of noisy and discrete datasets. TDA typically operates under the discrete setting where combinatorial structures such as graphs or simplicial complexes are imposed on the point cloud data to approximate their underlying structure. TDA, in our opinion, has over the past 15 years, brought a brand new perspective to topology in visualization.

The main data analysis tools in TDA are rooted in persistent homology [75], that is, the study of homology for point cloud data across multiple scales, and topological structures such as contour trees and Morse-Smale complexes. In the remainder of this section, we will discuss the applications of TDA for high-dimensional data visualization in the context of the visualization pipeline. For a complete taxonomy of topological methods in visualization including vector and tensor field visualization, see a recent survey by Heine et al. [79].

Many TDA techniques construct topological structures [88], [89] from scalar functions on point clouds (e.g., Morse-Smale complexes, contour trees, and Reeb graphs) as "summaries" over data. As a result, most TDA related techniques exist in the data transformation stage of the visualization pipeline. Among the commonly used TDA approaches, Reeb graphs/contour trees capture very different structural information of a real-valued function compared to the Morse-Smale complexes as the former is contour-based and the latter is gradient-based (Fig. 3). They both provide meaningful abstractions of high-dimensional data, which reduce the amount of data needed to be processed or stored; and they utilize sophisticated hierarchical representations that capture features at multiple scales, which enable progressive simplifications of features differentiating small- and large-scale structures in the data.

Morse-Smale Complex. The Morse-Smale complex (MSC) [90], [91] describes the topology of a function by clustering the points in the domain into regions of monotonic gradient flow, where each region is associated with a sink-source pair defined by local minima and maxima of the function. The MSC can be represented using a graph where the vertices are critical points, and the edges are the boundaries of areas with similar gradient behavior. The simplification of the MSC is obtained by removing pairs of vertices in the graph and updating connectivities among their neighboring vertices, thus merging nearby clusters by redirecting the gradient flow [92], [93], [94].

HDViz [5] employs an approximation of the MSC (in high dimensions) to analyze scalar functions on point cloud data. It creates a hierarchical segmentation of the data by clustering points based on their monotonic flow behavior, and designs new visual metaphors based on such a segmentation. This type of visual representation has been employed in the visual analytics of high-dimensional parameter spaces originating from simulations in nuclear engineering [4], [95], [96] and the National Ignition Campaign [97]. Correa and Lindstrom [98] suggest that by considering a different type of neighborhood structure, the accuracy in the extracted topology can be improved compared to those obtained within HDViz. The topological spine [99] uses the MSC to build an extremum graph that can more faithfully represent complex structures such as cycles and fractals occurring in the topology. Narayanan et al. [100] design a metric for comparing such extremum graphs of related data.

Reeb Graphs, Contour Trees, and Merge Trees. The Reeb graph of a real-valued function describes the connectivity of its level sets. A contour tree is a special case of the Reeb graph that arises in simply connected domains. A merge tree, also known as a barrier tree, is similar to Reeb graphs and contour trees except that it describes the connectivity of sublevel sets rather than level sets. The Reeb graph stores information regarding the number of components at any function value as well as how these components split and merge as the function value changes. Such an abstraction offers a global summary of the topology of the level sets and enables the development of compact and effective methods for modeling and visualizing scientific data, especially in high dimensions (i.e., [101], [102]). Approximating Reeb graphs from point cloud data are also possible [103]. For a more detailed history of the Reeb graph in computer graphics, see the survey by Biasotti et al. [104]

Mapper [102] decomposes data into a simplicial complex resembling a generalized Reeb graph and visualizes the data using a graph structure with varying node sizes. The software is shown to extract salient features in a study of diabetes by correctly classifying normal patients and patients with two causes of diabetes [105]. It is shown, in a restrictive sense, that Mapper converges to the Reeb space (a higher-dimensional generalization of Reeb graph) in the limit [106]. To this end, there have been some very recent efforts in understanding Reeb spaces and fiber surfaces via visualization, although those works have largely focused on bivariate functions on tetrahedral meshes [107], [108], [109]. Extensions of these works to general dimensionality are an open and interesting avenue of future research.

In terms of comparing the topologies of related data, the bottleneck distance between persistence diagrams is a well-established technique [110], but Beketayev et al. [111] have recently devised a more robust metric for comparing merge trees that accounts for the nesting structure of the tree.

Efficient algorithms for computing the contour tree [112], [113], [114], merge tree [115], and Reeb graph [116] in arbitrary dimensions have been developed. The latest state-of-the-art regarding contour trees have been parallel or distributed implementations, however these have focused specifically on tetrahedral meshes or regular grids in low dimensions [117], [118], [119], [120]. The visual representations of these topological structures are discussed in Section 4.4. For a more detailed reference of these methods in the time-varying setting, refer to the survey by Mascarenhas and Snoeyink [121].

Multi-Field Analysis. The methods mentioned above deal primarily with scalar field data (except for those regarding Reeb spaces [106], [107], [108], [109]), but more recently techniques have been developed to also deal with multi-field data. Jacobi sets [122] have been used to locate the critical points of one scalar field restricted to the level sets of another scalar field, allowing for the simultaneous comparison of two variables of interest. However, most applications of Jacobi sets have been to low-dimensional examples and are restricted to comparing only two outputs of interest.

A more recent and general technique is the development of the Joint Contour Net (JCN), a generalization of the Reeb graph introduced by Carr et al. [123], [124] that allows for the analysis of multi-field data. Duke and Hosseini [125] have subsequently improved the performance with a parallel implementation of the JCN, and Geng et al. have improved the interactivity by enabling brushing and linking and demonstrated its effectiveness in finding periodic patterns in oceanic data [126]. Chattopadhyay et al. [127], [128] have focused on bridging the gap between approximation and theory and produced an algorithm for performing simplification on the JCN among several other theoretical advancements.

The notion of Pareto optimality has also been explored. Pareto optimality is the trade-off analysis dealt with in multi-target optimization where a maximum implies increasing one target function value cannot be done without reducing another, and vice versa for a Pareto minimum. The simplicial Pareto set [129] builds off the technique proposed by Stadler and Flamm [130] to the piecewise linear setting in order to visualize the so-called Pareto sets of a sampled multivariate dataset. This work has been extended to deal with noisy data by using a reachability graph to perform topological simplification [131]. Huettenberger et al. [132] compare the JCN with the Pareto set and conclude that the JCN can be seen as a good and fast approximation of the Pareto set under specific conditions.

*Other Topological Features.*TDA also applies to nonfunctional data such as the focus of persistent homology where the connected components, circles, and voids in the data are studied. Carlsson [77] and Ghrist [78] both offer several applications of TDA and in particular highlight the topological theory used in a study of statistics of natural images [133]. Wang et al. [134] utilize TDA techniques developed by Silva et al. [135] to recover important structures in high-dimensional data containing nontrivial (high-dimensional branching and circular structures) topology. Rieck et al. utilize persistent homology to structurally compare high-dimensional datasets [136], [137] and to compare dimensionality reduction algorithms [138]. Bubenik [139] introduces a visualization called the persistence landscape as an alternative to the persistence diagrams and barcodes used by both Carlsson and Ghrist.

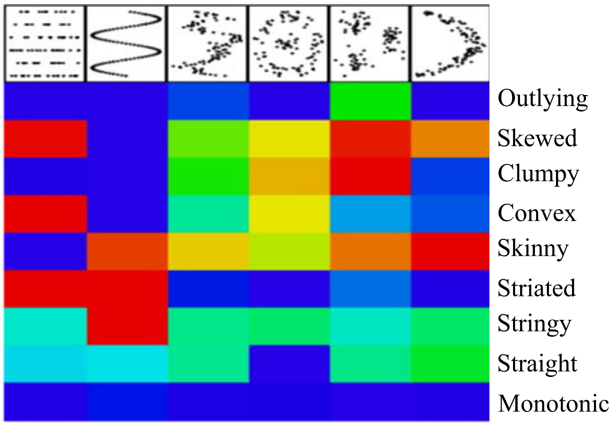


Fig. 4. Scagnostics introduced by Wilkinson et al. [140]

4 Visual Mapping

Visual mapping plays an essential role in converting the analysis result from the data transformation stage or the original dataset into visual structures for rendering in the view transformation stage. Based on differences in their structural patterns and visual compositions, we divide these approaches into axis-based, glyphs, pixel-oriented, hierarchy-based, and animation. Axis-based methods contain axes corresponding to the original data dimensions, projected dimensions, or combinations thereof. Glyphs encode information into the size, color, shape, and arrangement of small graphical symbols. Pixel-oriented techniques encode individual data values as pixels and focus on arranging the pixels in meaningful ways. Hierarchy-based mappings visualize nesting relationships in multiresolution and tree-like data. Animations include a temporal element to convey information in the changing of visual elements. In addition, the methods that evaluate the effectiveness of visual encodings are also discussed.

4.1 Axis-Based Methods

Axis-based methods refer to visual mappings where element relationships are expressed through axes representing the data dimensions. These methods include the most ubiquitous visual mapping approaches, such as scatterplot matrices (SPLOMs) and parallel coordinate plots (PCPs).

Scatterplot Matrix. A scatterplot matrix, or SPLOM, is a collection of bivariate scatterplots that allows users to view multiple bivariate relationships simultaneously. One of the primary drawbacks of SPLOMs is the scalability. The number of bivariate scatterplots increases quadratically with respect to the dataset's dimensionality. Numerous studies have introduced methods for improving the scalability of SPLOMs by automatically or semiautomatically identifying more interesting plots.

Originally introduced by John W. Tukey, Scagnostics are a set of measures designed for identifying interesting plots in a SPLOM. The recent works of Wilkinson et al. [140] extend the concept to include nine measures (illustrated in Fig. 4) capturing properties such as outliers, shape, trend, and density. In addition, they improve the computational efficiency by using graph-theoretic measures. Scagnostics have also been extended to handle time series data [141]. Guo [142] introduces an interactive feature selection method for finding interesting plots by evaluating the maximum conditional entropy of all possible axis-parallel scatterplots. The rank-by-feature framework [143] allows users to choose a ranking criterion, such as histogram distribution properties and correlation coefficients between axes, for scatterplots in SPLOMs.

Data class labels can play an important role in identifying interesting plots and selecting a meaningful ranking order. Sips et al. utilize class consistency [144] as a quality metric for 2D scatterplots. The class consistency measure is defined by the distance to the center of the class or entropies of the spatial distributions of classes. Tatu et al. [145] introduce different metrics for ranking the "interestingness" of scatterplots and PCPs for both classified and unclassified datasets. For data with labels, a class density measure and a histogram density measure are adopted as ranking functions for the scatterplots.

The ranking order provides only an indirect way to assess the scatterplots. Lehmann et al. [146] introduce a system for visually exploring all the plots as a whole. By reordering the rows and columns in the SPLOMs, this method groups relevant plots in the spatial vicinity of one another. In addition, an abstraction can be obtained from the reordered SPLOM to provide a global view.

Parallel Coordinates. Compared to a SPLOM, for which only bivariate relationships can be directly expressed, the parallel coordinate plot (PCP) [6], [7], [147] allows patterns that highlight multivariate relations to be revealed by showing all the axes at once. For a given n-dimensional dataset, theoretically, there are $n!$ permutations of the ordering of the axes. With different axes order, vastly different information may be presented. Therefore, one of the fundamental challenges when dealing with PCPs is determining the appropriate orders of the axes [7]. Since a user typically can only interpret the visual patterns among nearby axes, the search space can be drastically reduced by focusing on localized axes orders, such as consecutive dimension triples (an axes and its immediate neighbors) or pairwise dimensions. For these scenarios, finding the minimum number of permutations needed to display all dimension triples or pairwise dimension combinations is the goal. Hurley et al. [148] adopt Eulerian tours and Hamiltonian decompositions of complete graphs to generate axis order permutations ($O(n/2)$) covering all bivariate patterns between dimensions. Inselberg has posed the problem of finding permutations that display all adjacent triples [6], which may be considered as a future visualization challenge in PCPs.

A few other methods utilize quality metrics and subspace finding methods to automatically identify interesting axes orders. The PCP ranking methods developed by Tatu et al. [145] work for both classified and unclassified datasets. For unlabeled data, the Hough space measure is used, and for labeled data, a similarity measure and overlap measures are adopted. Ferdosi et al. introduce a dimension ordering method [149] that is applicable for both PCPs and SPLOMs utilizing the subspace analysis method from their earlier work [59] discussed in Section 3.3. Johansson and Johansson [150] propose an interactive system adopting a weighted combination of quality metrics for dimension selection and automatic ordering of the axes to enhance visual patterns such as clustering and correlation.

In addition, as the number of data points increases, the line density in the PCP increases dramatically, which can lead to visual clutter [7] thus hindering the discovery of patterns (e.g., density variation, dimension correlation). As a result, clutter reduction through filtering, aggregation, visual encoding, and dimension reordering, is another important challenge for PCPs. Interactive filtering of data, such as brushing linked axes, is essential for alleviating visual clutter. Chapter 10 of Inselberg's book [6] provides a great discussion on how to exploit interactivity in PCPs to understand large and complex data. A set of query operations, which can be combined to construct more complex queries, is identified as the basis for the exploration.

Aggregation and visual encoding can also be used in combination with interactive exploration to reduce visual clutter. In the work by Novotny and Hauser [151], a focus+context visualization scheme is adopted for reducing the clutter by aggregation. In this approach, the outliers are indicated by single lines and the trends that capture the overall relationship between axes are approximated by polygon strips. Zhou et al. introduce a line bundling scheme [152] for enhancing the visual clusters. The authors exploit the curved edges and arrange the edges by minimizing the curvature while maximizing the parallelism of the adjacent ones. The progressive parallel coordinate (PPC) [153] work introduces several LOD-hierarchy based visual encoding approaches to address the challenges of large datasets and overplotting. In the work introduced by Dang et al. [154], density is expressed by stacking overlapping elements. For the PCP case, a 3D visualization is presented, where either the edges are stacked as curves or the points on the axes are stacked vertically as dots to alleviate the clutter with an additional dimension. Finally, as dimension ordering can greatly affect the PCPs' expressiveness, Peng et al. [155] introduce a clutter reduction method for PCPs by reordering the axes. Clutter reduction methods that employ screen space measures are discussed in detail in Section 5.4.

Radial Layout. The star coordinate plot [156], also referred to as a bi-plot [157], is a generalization of the axis-aligned bivariate scatterplot. The star coordinate axes represent the unit basis vectors of an affine projection. The user is allowed to modify the orientation and the length of the axes as a way of altering the projection. However, due to the unbounded manipulation, star coordinates may produce affine projections in which substantial distortion occurs. Lehmann et al. extend the star coordinate concept with an orthographic constraint [158], which better preserves the structure of the original dataset in the projection.

Radviz [157], similar to the star coordinates, adopts a circular pattern. The difference is that Radviz does not define an explicit projection matrix. In Radviz, n-dimensional anchors are placed along the perimeter of a circle, each representing one of the dimensions of an n-dimensional dataset. A spring model is constructed for each point, where one end of a spring is attached to a dimensional anchor and the other is attached to the data point. The point is then displayed where the sum of the spring forces equals zero. Albuquerque et al. [159] devise a RadViz quality measure allowing automatic optimization of the dimensional anchor layout.

DataMeadow [160] introduces a radial visual encoding named DataRoses, which is represented as a PCP laid out radially as opposed to linearly. Lastly, PolarEyez [161] introduces a focus+context visualization in which the high-dimensional function parameter space is encoded in a radial fashion around a user-controlled focal point. Data near the focal point is represented with more precision, and the focal point can be altered to focus on different parts of the data.

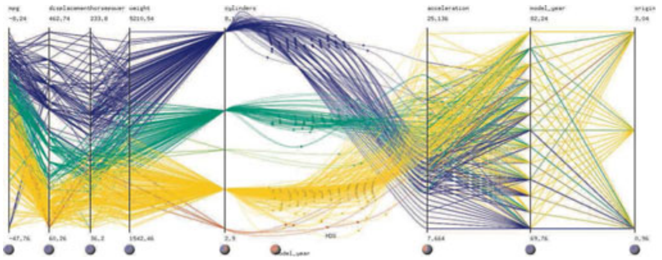


Fig. 5. Scattering points in parallel coordinates by Yuan et al. [162].

Hybrid Construction. The axis-based methods can also be combined to create new visualizations. The scattering points in parallel coordinate work [162] (Fig. 5) embeds an MDS plot between a pair of PCP axes. The flexible linked axes work [163] is a generalization of the PCP and the SPLOM. The tool gives the user the ability to create new configurations by drawing and linking axes in either scatterplot or PCP style. Proposed by Fanea et al., the integration of parallel coordinate and star glyphs [164] provides a way to "unfold" the overlapped values in the PCP axis in 3D space. In this work, each axis in the PCP is replaced by a star glyph that represents the values of the corresponding dimension across all points, and then each high-dimensional point is described as a set of line segments in 3D connecting the individual values in the star glyphs.

In addition, a number of visual representations derive from the well-known visual encodings. Angular histograms [165] introduced a novel visual representation that improves the scalability of PCPs by summarizing the trend of the line segments between the axes. The tiled PCP [166] adopts a row-column 2D configuration instead of the 1D linear layout of the traditional PCP for simultaneous visualization of multiple time steps and variables.

4.2 Glyphs

By rendering "small graphical symbols", the glyphs-based approaches utilize shape, color, opacity, size, location, etc. to encode high-dimensional information.

Chernoff faces [167] are one of the first attempts to map a high-dimensional data point into a single glyph. The system works by mapping different facial features to separate dimensions. In a few recent works, glyphs have been utilized to provide statistical and sensitivity information in order to present trends in the data. By utilizing local linear regression to compute partial derivatives around sampled data points and representing the information in terms of glyph shape, sensitivity information (uncertainty related topics are discussed in Section 7.1) can be visually encoded into scatterplots [168], [169], [170], [171].

The methods described above deal with encoding per data point information into glyphs. Other usages of glyphs attempt to show the trends in parts of the data. DICON [172] uses dynamic icons based on treemap visualization to encode clusters of data into separate glyphs, and allows the user to interactively merge, split, filter, regroup, and highlight information within clusters. Lehmann et al. [173] introduce *visualnostics*, in which various 2D representations of high-dimensional data such as parallel coordinates, scatterplots, RadViz, and star coordinates are summarized by pictograms to aid visual search tasks.

Finally, Ward [174] gives a thorough, practical treatment for generating and organizing effective glyphs for multivariate data, paying particular attention to the common pitfalls involving the use of glyphs.

4.3 Pixel-Oriented Approaches

In an effort to encode the maximal amount of information, several works have targeted dense pixel displays. Researchers have focused on encoding data values as individual pixels and creating separate displays, or *subwindows*, for each dimension.

Some of the earliest works in this area date back to the mid 1990s [175], [176]. VisDB [175] visualizes database queries by creating a 2D image for each dimension involved in the query and mapping individual values of a dimension to pixels. The mapped data is sorted and colored by relevance such that the data most related to the query appears in the center of the image, and the data spirals outward as it loses relevance to the query. Circle segments [176] arrange multidimensional data in a radial fashion with equal size sectors being carved out for each dimension.

The pixel concept can be applied to bar charts to create pixel bar charts [177]. Pixel bar charts first separate data into separate bars based on one dimension or attribute, and they can also split the data along the orthogonal direction using another dimension, although most results are reported using only one direction for splitting data. Once split, the data points are sorted along the horizontal axis within the bars using one dimension and ordered along the vertical axis using another dimension. Wattenberg introduces the jigsaw map [178], which again maps data points to pixels and uses discrete space-filling curves in order to fill a 2D plane in a more sensible fashion than a comparative treemap layout.

The Value and Relation (VaR) displays [179] combine the recursive pattern displays [180] with MDS in order to lay out the separate subwindows such that similar dimensions are placed closer together. A latter iteration [181] enhances the work by providing alternative dimension representations and their layout schemes.

4.4 Hierarchy-Based Approaches

Hierarchical structures can be used to capture dimensional relationships and to provide summaries for representing high-dimensional datasets.

Dimension Hierarchies. Large numbers of dimensions hinder our ability to navigate the data space and cause scalability issues for visual mapping. A hierarchical organization of dimensions explicitly reveals the dimension relationships, helping to alleviate the complexity of the dataset. Yang et al. propose an interactive hierarchical dimension ordering, spacing, and filtering approach [182] based on dimension similarity. The dimension hierarchy is represented and navigated by a multiple ring structure (InterRing [183]), where the innermost ring represents the coarsest level in the hierarchy.

Topological Hierarchies. In the previous section, we have discussed topological structures, which can provide a ranking of features with the help of persistence simplification and thus be treated as a hierarchy.

The contour tree that summarizes the structure of (potentially) high-dimensional data has been the subject of many visual manifestations with a focus on its 2D graph drawing; Heine et al. establish a set of constraints to produce aesthetic and interpretable visualizations of this nature [184]. More abstract visual metaphors have been introduced, such as orreries [185], cacti [186], and landscapes [187], [188], [189], [190], [191], [192]. These visual metaphors can be and have been used to support high-dimensional data visualization by abstracting the structures in high dimensions as a low-dimensional representation, where its layout is used to convey the hierarchy and proximity of features. In particular, Weber et al. [192] have presented such a metaphor for visually mapping the contour tree of high-dimensional functions to a 2D terrain. The metaphor preserves the relative size, volume, and nesting of the topological features. Harvey and Wang [189] have extended this work by computing all possible planar landscapes. They are able to preserve exactly the volumes of the high-dimensional features in the areas of the terrain. In addition, the works of Oesterling et al. [190], [191] have used this same metaphor to visualize a related structure, the join tree. They use a novel high-dimensional interpolation scheme in order to estimate the density from the raw data points and visually map the density as points on top of their generated terrains. Oesterling et al. [193] have continued this line of work by creating a linked view software system including user interactions in the analysis by allowing users to brush and link with PCPs and PCA projections of the data. In addition, they have presented a new method of sorting the features based on persistence, cluster size, or cluster stability, thus adjusting the placement of features in the topological landscape. The level set tree proposed by Klemelä [194] is a similar data structure to the contour tree used in understanding multivariate density distributions as piecewise constant functions. Klemelä provides three visualizations for understanding the statistical and shape properties of the distributions: a tree drawing, a barycenter plot, and a volume plot.

In terms of visual mapping for Morse-Smale complexes, skeletons are often used to convey their topology; however, these may not be the best visualization technique, particularly in the face of uncertainty. An alternative is to apply a graph-based layout (i.e., [195], [196], [197]) to the MSC, and combine such a layout with dimension reduction and statistical techniques such as regression to produce content-rich visual representations, e.g., HDViz [5].

Other Hierarchical Structures. In the structure-based brushes work [198], a data hierarchy is constructed to be visualized by both a PCP and a treemap [199], allowing users to navigate among different levels-of-detail and select the feature(s) of interest. The structure decomposition tree [200] presents a novel technique that embeds a cluster hierarchy in a dimensional anchor-based visualization using a weighted linear dimension reduction technique. It provides a detail plus overview structural representation and conveys coordinate value information in the same construction. The system supports user-guided pruning, optimization of the decision tree, and encoding the tree structure in an explorable visual hierarchy. Kreuseler et al. present a novel visualization technique [201] for visualizing complex hierarchical graphs in a focus+context manner for visual data mining tasks.

4.5 Animation

As stated in Heer et al.'s work [202], animation, when used appropriately, can significantly improve graphical perception. Many techniques for visualizing high-dimensional data utilize animated transitions to enhance the perception of point and structure correspondences among multiple relevant plots.

The GGobi system [203] provides a mechanism for calculating a continuous linear projection transition between a pair of linear projections based on the principal angles between them. In the Rolling the Dice work [204], a transition between any pair of scatterplots in a SPLOM is made possible by connecting a series of 3D transitions between scatterplots that share an axis. RnavGraph [205] constructs a graph connecting a number of interesting scatterplots. A smooth animation is generated between all scatterplots that are connected by an edge. The $TripAdvisor^{ND}$ [56] system allows users to explore the neighborhood of a subspace by tilting the projection plane using a polygonal touchpad interface.

4.6 Perception Evaluation

The design goal of visual mapping and encoding is to directly convey the information to the user through visual perception. The evaluation of this mapping is vitally important in determining the effectiveness of the overall visualization.

Sedlmair et al. have carried out an extensive investigation of the effectiveness of visual encoding choices [206], including 2D scatterplots, interactive 3D scatterplots, and SPLOMs. Their findings reveal that the 2D scatterplot is often decent, and certain dimension reduction techniques provide a good alternative. In addition, SPLOMs sometimes add additional value, and the interactive 3D scatterplot rarely helps and often hurts the perception of class separation. A perception-based evaluation [207] of various projection methods that generate 2D linear or nonlinear scatterplot is presented by Etemadpour et al. In this work, the authors identify eight typical tasks that relate to the properties of projection methods and results in terms of segregation capability, projection precision, and incurred visual cluttering. The evaluation demonstrates that the projection performance is task dependent and heavily depends on the nature of the data. In addition, certain projections perform better on specific types of tasks.

The efficacy of several PCP variants for cluster identification has been studied in [208]. A comparison is performed among nine PCP variations based on existing methods and combinations of them. The evaluation reveals that, aside from the scatterplots embedded into parallel coordinates, a number of seemingly valid improvements do not result in significant performance gains for cluster identification tasks. A comparative study between two popular radial visualizations, the RadViz and star coordinates, can be found in [209]. As pointed out in the study, RadViz is useful for analyzing sparse data, but the nonlinear nature of its normalization step impedes its application and accuracy compared to the flexible and linear star coordinates. An evaluation of radial visualization solutions for composite indicators (a measuring and benchmark tool used to capture multidimensional concepts) is presented by Albo et al. [210]. Heer et al. investigate the animated transition effectiveness between statistical graphs [202], such as bar charts, pie charts, and scatterplots. Their results reveal that animated transitions, when used appropriately, can significantly improve graphical perception.

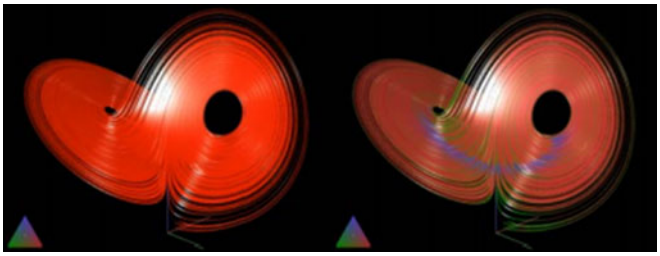


Fig. 6. Illuminated 3D scatterplot by Sanftmann et al. [211].

5 View Transformation

View transformations dictate what we ultimately see on the screen. As pointed out by Bertini et al. [8], the view transformation can also be described as the rendering process that generates images in the screen space.

5.1 Illustrative Rendering

Illustrative rendering describes methods aimed at achieving a specific visual style by applying custom rendering algorithms. The illustrative PCPs work [212] provides a set of artistic rendering techniques for enhancing visual patterns (e.g., line density) in PCPs. Illuminated scatterplots [211] (Fig. 6) classify points based on the eigenanalysis of the covariance matrix and give the user the opportunity to see effects such as planarity and linearity when visualizing dense scatterplots. Johansson et al. [213] reveal structures in PCPs by adopting the transfer function concept commonly used in volume rendering. Based on user input, the transfer function maps the line densities into different opacities to highlight different features.

Illustrative renderings are also used for highlighting focal areas, such as the well-known TableLens approach [214] for visualizing large tables. Such a magic lens based approach permits fast exploration of an area of interest without presenting all the details and, therefore, reduces clutter in the view. MoleView [215], for visualizing scatterplots and graphs, adopts a semantic lens for allowing users to focus on the area of interest and keep the in-focused data unchanged while simplifying or deforming the rest of the data to maintain context. A survey on early distortion-oriented magic lens techniques is presented by Leung and Apperley [216].

5.2 Continuous Visual Representation

For most high-dimensional visualization techniques, a discrete visual representation is assumed since each element usually corresponds to a single data point. However, due to limitations such as visual clutter and computational cost, many applications prefer a continuous representation.

The work of Bachthaler and Weiskopf [217] presents a mathematical model for constructing a continuous scatterplot. The follow-up work [218] introduces an adaptive rendering extension for continuous scatterplots, thereby increasing the rendering efficiency. This concept is extended to create continuous PCPs [219] based on the point-line duality between scatterplots and parallel coordinates. In addition, Lehmann et al. introduce a feature detection algorithm designed for continuous PCPs [220].

Clutter in PCPs and scatterplots leads to occlusion of data distribution patterns. In the splatterplot work [221], the authors introduce a hybrid representation for scatterplots to overcome the overdraw issue when scaling to very large datasets. The proposed abstraction automatically groups dense regions into an abstract contour and renders the rest of the area using selected representatives, thus preserving the visual cue for outliers. A splatting framework for extracting clusters in PCPs is presented in [222], where a *polyline splatter* is introduced for cluster detection, and a *segment splatter* is used for clutter reduction.

5.3 Accurate Color Blending

When rendering semitransparent objects, color blending methods have a significant impact on the perception of order and structure. As stated in the hue-preserving color-blending work [223], the commonly adopted alpha-compositing can result in false colors that may lead to a deceiving visualization. The authors propose a data-driven machine learning model for optimizing and predicting hue-preserving blending. This model can be applied to high-dimensional visualization techniques such as illustrative PCPs [212], where a depth ordering clue is better preserved. In the Weaving versus Blending work [224], the authors investigate the effectiveness of two color mixing schemes: color blending and color weaving (interleaved pattern). The results indicate that color weaving allows users to better infer the value of individual components; however, as the number of components increases, the advantage of color weaving diminishes.

5.4 Image Space Metrics

As discussed in Section 4.1, a number of quality measures have been proposed to analyze the visual structure and automatically identify interesting patterns in PCPs or scatterplots. In this section, we discuss the image space based quality measures that are applied in the screen space.

Arterode et al. propose a method [225] for uncovering clusters and reducing clutter by analyzing the density or frequency of the plot. Image processing based techniques such as grayscale manipulation and thresholding are used to achieve the desired visualization. Johansson et al. introduce a screen space quality measure for clutter reduction [226]. The metric is based on distance transformation, and the computation is carried out on the GPU for interactive performance.

Pargnostics [227], a portmanteau for parallel coordinates and diagnostics (similar to Scagnostics [140]), is a set of screen space measures for identifying distinct patterns among pairs of axes in PCPs. The metrics include line crossings, crossing angles, convergence, and overplotting. For each metric, the system provides ranked views for pairs of axes, allowing the user to guide exploration and visualization. Pixnostic [228] is an image space based quality metric

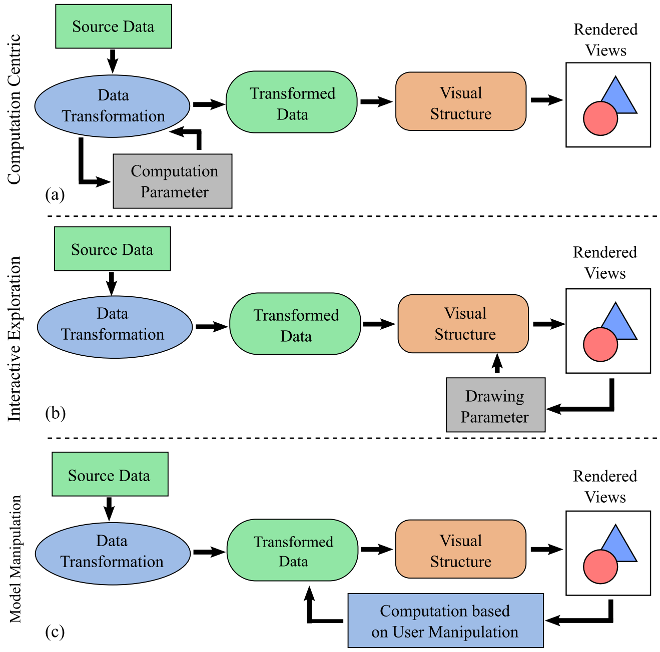


Fig. 7. The three types of user interaction paradigms with varying degrees of user involvement. Since each paradigm can interact with each processing stage in the visualization pipeline, the diagram highlights the most general patterns.

for ranking interestingness for pixel-based (Section 4.3) visualization such as Pixel Bar Chars [177].

6 User Interaction

As illustrated in Fig. 1, interaction is integrated with each processing stage. In this section, we identify three types of user interaction (computation-centric approaches, interactive exploration, and model manipulation) based on the amount and type of user involvement and illustrate how they interact with the different stages of the visualization pipeline (see Fig. 7). In both recent surveys [229], [230] on user interaction in visualization applications, the level of integration between the computation and visualization (including user interaction) is used for classifying the methods. In many ways, their classifications are aligned with the proposed approach, with the distinction that our discussion is directly linked to the visualization pipeline. In the following sections, we will discuss each paradigm in detail.

6.1 Computation-Centric Approaches

Computation-centric (see Fig. 7a) approaches require only limited user input such as setting initial parameters. These methods center around algorithms designed for well-defined computational problems such as dimension reduction [31], [33], [36], [37], subspace clustering [55], [57], [59], [64], regression analysis [65], [68], quality metric based ranking [140], [145], etc. Computation-centric approaches are most concentrated in the data transformation stage (as illustrated in Fig. 7a).

6.2 Interactive Exploration

Interactive exploration (see Fig. 7b) approaches navigate, query, and filter the existing model interactively for more effective visual communication. They mostly exist in the visual mapping stage, where the visual structure is interactively modified by user interaction. The distinction between interactive exploration and model manipulation (see Fig. 7c) is made to highlight the fact that users do not alter the underlying computation model in the interactive exploration.

In the data transformation stage, the interactive exploration scheme allows users to guide progressive dimension reduction, where a partial result is presented upon request [36]. In the works by Turkay et al. [2], [52] and Yuan et al. [53], a subset of dimensions is interactively selected and explored in dimension space. In the visual mapping stage, a large number of methods focus on interactive visual exploration through filtering, zooming, distorting, linking, and brushing of visual representations. For example, recent works [231], [232] by Gratzl et al. introduce interesting interactive methods for ranking multiple attributes and exploring subsets of tabular datasets. Interactive exploration methods also play an important role in the Knowledge Discovery in Databases (KDD) and data mining process, where the term *visual data mining* [11], [12], [233] is introduced (see Section 7.2). In the view transformation stage, interactivity mostly originates from the changing of rendering parameters and configurations, which appears in both the magic lens based methods [215], [216] and the illuminated 3D scatterplots [211] (discussed in Section 5.1).

6.3 Model Manipulation

Model manipulation (see Fig. 7c) techniques represent a class of methods that integrate user manipulation as part of the algorithm and update the underlying model to reflect the user input to obtain new insights.

Take the distance function learning work [43], for example. The initial embedding is created using a default distance measure. Through interaction, the initial point layout is modified based on the expert user's domain knowledge. The system then adjusts the underlying distance model to reflect the user input. Such a process is illustrated in Fig. 7c. Hu et al. present a method [234] for improving the translation of user interaction to algorithm input (visual to parameter interaction) for distance learning scenarios. Explainers [44] are projection functions created from a set of user-defined annotations. Similarly, in recent work [235], Kim et al. introduce an approach for steering axis-aligned linear projections by dragging points into x or y axes to generate new linear projections that reflect the combination of data attributes bound to the axes. The control point based projection methods [40], [41], [42] update the overall projection result based on user manipulation of the control points. Liu et al. [49] introduce a projection manipulation scheme facilitates the understanding of high-dimensional data via direct modification of its 2D embedding. Distortion metrics are used for feedback during the manipulation.

7 Emerging Areas

In this section, we identify a couple of emerging areas that could inspire future research in high-dimensional data visualization. However, due the subjective nature of such a discussion, our intention is not to give a comprehensive review of all possible future directions, but rather to describe specific directions with adequate details.

7.1 Uncertainty in High Dimensional Data Visualization

Along with the large scale and high dimensionality of the data, information pertaining to uncertainty is becoming increasingly available and important.

Uncertainty visualization has been deemed a top research problem in scientific visualization [236], due to the increasing availability of uncertainty information from simulation and the importance of understanding data quality, confidence, and error issues when interpreting scientific results. Visualizing the uncertainty in data and the examination of uncertainty in the visualization pipeline are also essential for high-dimensional data visualization.

Similar to surveys on the topic [237], [238], [239], [240], we make a distinction about the source of uncertainty. In the first case, the process of acquiring the data imposes uncertainty that must be communicated to the user. In the second case, the transformations the data undergoes before appearing onscreen can also add uncertainty. We denote the prior case as *data uncertainty* and the latter case as *algorithmic uncertainty*.

*Data Uncertainty.*When the data is encoded with its own inherent uncertainty or the goal is to summarize an ensemble of data, this extra information must be visually encoded. Typical techniques include blurring of visual marks [241], [242], [243], glyphs [244], or colormaps and noise [245], [246] to indicate ranges of uncertainty. However, such a simplistic treatment of uncertainty often causes problems in data understanding. By increasing the illegibility or complexity of an image corresponding to the amount of trustworthiness of the data, the amount of coherent information from that image decreases, which leads to less usable information within a visualization. In contrast, more recent works attempt to express a more visually quantifiable encoding of uncertainty by using summaries of the data to reduce the visual clutter [247], [248], [249]. For example, Chen et al. [247] perform uncertainty-aware dimensionality reduction on ensemble data by accounting for the distribution of the ensemble. In the numerical weather model ensemble visualization work [249], Sanyal et al. replace the traditional spaghetti plots (line plot representation for each element in the ensemble) with a combination of visual elements, including ribbons and glyphs, that quantify the uncertainty by summarizing individual ensemble member's standard deviation, interquartile range, and the confidence interval. Limited work exists that specifically targets high-dimensional data, which is why we believe the extensions and generalizations of existing uncertainty visualization capabilities (e.g., [246], [248], [249]) to high-dimensional data are important future directions.

*Algorithmic Uncertainty.*Another interesting aspect of uncertainty quantification is based on uncertainty introduced in the visualization pipeline (shown in Fig. 1). The concept of uncertainty-aware visual analytics is first discussed by Correa et al. [168]. In this work [168], the authors measure the uncertainty introduced by three common data transformation techniques, namely regression, principal component analysis, and k-mean clustering. Similar concepts are further explored by other works [48], [49], [51], [250], where the uncertainty (e.g., bias and distortions) stemming from the dimension reduction is quantified and visualized. In addition, other examples targeting high-dimensional data visualization have focused on analyzing the uncertainty with respect to the accuracy of a fitted model (see Section 3.4 and [251] for more details). These methods mostly focus on the uncertainty stemming from the data transformation stage. However, more work can be done to define measures of uncertainty associated with the two latter processing stages in the visualization pipeline: visual mapping and view transformation.

7.2 The Interplay Between Data Science and High Dimensional Data Visualization

Data science is an interdisciplinary area where multiple subjects are brought together. By relying on solid foundations in mathematics and statistics, and the effective tools from computer science, data science aims at transferring data into knowledge for solving real world problems. Visualization as an integral part of data science plays an important role in the data analysis process. In the following sections, we will look into several aspects of data science and discuss their connection with high-dimensional data visualization and possible emerging research directions.

Data Management.In the visualization literature, data management is often considered as an optional subsystem, which is rarely the focus of the study. However, the increasing complexity and size of the data and demands for data-centric analysis call for robust and flexible data management systems. An interesting integration of data management and visualization system can be found in the VisTrails framework [252]. VisTrails manages the data and metadata of visualization results, and provides the ability to trace and compare the history of different visualization pipeline configurations, which allows for fast exploration and discovery. A relational database is usually adopted for managing data for visualization. However, its limited query efficiency for high-dimensional data leads to the introduction of more efficient index schemes such as the X-tree [253], which is designed for high-dimensional data.

Besides aiding the visualization process with the integration of databases, visualization can also be adopted as an intuitive interface for querying databases. Such an interface can translate the user intention into database queries and then present the results in visual forms. The increasingly sophisticated interplay between visualization and database querying tasks leads to the introduction of visual data mining and related techniques discussed below.

Data Mining. Data mining studies the process of extracting meaningful patterns or relationships from data by utilizing various statistical or machine learning algorithms and efficient data management infrastructures. Many purely automatic, analytical approaches have been introduced and produce reasonable results. However, especially in the recent years, our ability to generate, collect, and store data has quickly outweighed our ability to analyze it. One important paradigm shift in addressing challenges comes from the realization that for resolving a complex analytical problem, the involvement of humans in the early stage of the analysis process is crucial. Instead of relying solely on confirmatory data analysis, exploratory data analysis [254] and visual data exploration [12] have proven to be extremely valuable and effective. Visual data mining [11], [233] is one outcome of such development. It bridges the gap between visual data exploration and data mining tasks. It not only provides a more intuitive interface for communicating the underlying computational model to the user, but also exploits the human vision system for pattern searching to deal with the ever-increasing size and complexity of data. Such a paradigm is currently recognized as part of the emerging visual analytics [255], [256] field, which is described as the science of analytical reasoning supported by interactive visual interfaces.

Inspired by these new techniques, many high-dimensional data visualization techniques combine automatic analysis with user-driven visual exploration. Stolte et al. introduce Polaris [257], which is a visual query and analysis system designed for relational databases. In this system, relational queries can be defined by visual specifications that allow fast incremental development and intuitive understanding of the data. The authors later extend their work for hierarchically structured data cubes [258]. In their last installment [259], a multiscale visualization system utilizing Polaris and the data cubes extension is introduced. The Polaris system was later developed into the well-known commercial visualization system Tableau. Hao et al. introduce the Intelligent Visual Analytics Queries [260]. Their approach utilizes correlation and similarity measurements that are then encoded by summary visualization for mining localized data relationships. Detailed surveys and discussions on the topic of visual data mining can be found in [10], [11], [12]. We believe new research can stem from the further development and interaction among data mining tasks, visual encoding and exploration, and in-depth user interactions with the full spectrum of the analysis process.

Machine Learning. Machine learning introduces the tools to build models from data for predicting or summarizing unknown data. It provides building blocks for constructing higher-level tasks commonly found in data mining and artificial intelligence. Machine learning algorithms such as manifold learning [32] and subspace clustering [61], [261] have been adopted for visualizing high-dimensional data.

On the other hand, the high-dimensional visualization methods also aid intuitive understanding of the algorithm and the parameter tuning process. The fundamental tasks of machine learning involve the study of the feature space and the learned models from the data, which are high-dimensional in nature. However, intuitive understanding and exploration of these high-dimensional models are extremely difficult. To resolve such a challenge, several visualization approaches have been introduced to provide visual aids. Tzeng et al. present a visualization system that helps users design neural networks more efficiently [262]. The works of Teoh and Ma [263] and van den Elzen and van Wijk [264] investigate visualization methods for interactively constructing and analyzing decision trees. Visualization has also been used to aid model validation [265], [266] (regression model related validation and tuning is discussed in more detail in Section 3.4). Garg et al. use Hidden Markov Models as an example to illustrate the effectiveness of their visualization approach [267]. It achieves a balance between manual operation and a fully automatic approach for tasks such as data tagging by involving the user in the decision-making process.

Numerous challenges for understanding machine learning algorithms coincide with the goal of high-dimensional visualization. We believe high-dimensional visualization will play an increasingly important role in designing, tuning, and validating machine learning algorithms. At the same time, more machine learning algorithms will also find their way into visualization methods.

8 Summary and Reflections

In this survey, we aim to provide a structured overview of distinct subfields, in which new methods can be inspired based upon combinations and extensions of existing approaches. To better achieve this goal, in this section, we summarize and reflect on each stage of the visualization pipeline and focus on the scenarios in which various methods can be effectively applied.

Starting from the data transformation stage, the methods in this stage of the pipeline are computation-centric and mostly focus on obtaining quantitative results. Dimension reduction methods are commonly used for capturing the overall structure of a dataset. Since these methods are designed for reducing the dimension while preserving the important structures, dimension reduction approaches are more suitable for handling data with a large number of dimensions compared to many visual mapping approaches (e.g., scatterplot matrix). The scalability of dimension reduction methods as visualization tools is addressed partly by the development of the control point based projection approaches [40], [41], [42] and partly by approximations [36], [37], [38]. Various precision measures [46], [47], [49] have been introduced to provide a per-point assessment of the accuracy in terms of information preservation, which is essential for interpreting the results.

Due to the complexity of high-dimensional data, it is unlikely a single embedding (produced by dimension reduction) is sufficient for understanding every dataset. Instead, identifying multiple informative 2D projections automatically or semiautomatically is essential for exploring different aspects of the data. The subspace clustering methods either find clusters in subset of the dimensions (originated from data mining [58]) or cluster points that share a low-dimensional linear subspace (originated from machine learning [61]). These methods not only help in identifying multiple interesting projections but also address the challenges of the ever-increasing complexity of the data (e.g., number of dimensions) by dividing them into lower dimensional subsets.

Besides the approaches focusing on generating one or multiple low-dimensional embeddings, regression analysis provides a class of methods designed to capture the quantitative relationship among individual dimensions. Interactive visualization has been integrated with the regression analysis process for more effective parameter exploration and tuning [65], [66]. Finally, topological data analysis (Section 3.5) provides a unique approach for summarizing high-dimensional structure, which we believe will play an increasingly important role in high-dimensional data visualization.

The next stage of the pipeline is visual mapping. The most common visual representations for high-dimensional data, such as SPLOMs and PCPs, are built around the different arrangements of data axes. A SPLOM helps capture the complete bivariate relationship by permuting all possible pairs of the axis whereas a PCP [6], [147] provides a single view of the multivariate relationship by showing all the axes vertically. The major drawback of both approaches is that the number of possible 2D configurations increases drastically as the dimensionality increases. As a result, various quality metrics [140], [145], [268] that help automatically filter for the interesting configurations are at the center of recent developments. Other axis configurations, such as radial layouts, are also gaining popularity [159], [160], [161]. In addition, recent advances have introduced new visual representations by coupling existing approaches to combine the advantages of different visual representations [162], [163], [164].

Glyph-based approaches [167], [168], [170], [172] are among the earliest methods for visualizing high-dimensional data. They either encode and highlight certain per-point information or combine multiple points to express summary information. The pixel-oriented representations [176], [177], [178], [181] are closely related to the glyphs, but instead of encoding individual data points, they mostly provide a compact representation of dimensions that are packed as pixels. Hierarchy-based representations [182], [200] are usually a natural translation of a structure that is tree-like or multiresolution in nature, such as encoding the dimension hierarchy [182] and the hierarchical topological segmentation [185], [187]. In the past decades, many works have focused on evaluating the effectiveness of various visual encodings, such as PCPs [202] and SPLOMs [206]. The perception of visual effects such as animation [56], [202], [204] has also been studied. Such development highlights the important trend where rigorous evaluation is an integral part of any effective visualizations.

The last stage of the pipeline is the view transformation, which describes the process of generating rendered images from visual structures. Many innovative methods in this stage focus on enhancing the existing rendering techniques to address their limitations or highlight the regions of interest. The illustrative rendering works [212] aim at emphasizing certain aspects of the data through visual exaggeration while discarding other less important visual properties. The continuous visual representations [217], [218] are designed to address overplotting issues through analytical modeling and splattering approximations. The techniques [212], [223] that address color blending have a similar goal. Instead of the general overplotting issue, they focus on resolving the challenge of misleading overlapping colors. In addition, the image space metrics [225], [226], [227] are a natural extension from the quality metrics of visual structures (discussed in Section 4), for which evaluating the metric is more efficient in the image space (usually for dealing with a large number of points).

Finally, we discuss the emerging areas in high dimensional data visualization, namely uncertainty quantification (Section 7.1) and data science (Section 7.2). We believe the interaction between these topics and high-dimensional data visualization will lead to many interesting future research and applications.

Acknowledgments

This work was performed in part under the auspices of the US DOE by LLNL under Contract DE-AC52-07NA27344., LLNL-CONF-658933. This work is also supported in part by US National Science Foundation IIS-1513616, US National Science Foundation 0904631, DE-EE0004449, DE-NA0002375, DE-SC0007446, DE-SC0010498, NSG IIS-1045032, NSF EFT ACI-0906379, DOE/NEUP 120341, DOE/Codesign P01180734.

REFERENCES

[1] R. Clarke, et al., "The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data," *Nature Rev. Cancer*, vol. 8, no. 1, pp. 37–49, 2008.

[2] C. Turkay, P. Filzmoser, and H. Hauser, "Brushing dimensions—A dual visual analysis model for high-dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2591–2599, Dec. 2011.

[3] D. Engel, et al., "Visual steering and verification of mass spectrometry data factorization in air quality research," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2275–2284, Dec. 2012.

[4] D. Maljovec, B. Wang, V. Pascucci, P.-T. Bremer, and D. Mandelli, "Analyzing dynamic probabilistic risk assessment data through topology-based clustering," in *Proc. Int. Topical Meeting Probabilistic Safety Assessment Anal.*, 2013, pp. 1839–1854.

[5] S. Gerber, P. Bremer, V. Pascucci, and R. Whitaker, "Visual exploration of high dimensional scalar functions," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1271–1280, Nov./Dec. 2010.

[6] A. Inselberg, *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Berlin, Germany: Springer, 2009.

[7] J. Heinrich and D. Weiskopf, "State of the art of parallel coordinates," *STAR Proc. Eurographics*, vol. 2013, pp. 95–116, 2013.

[8] E. Bertini, A. Tatu, and D. Keim, "Quality metrics in high-dimensional data visualization: An overview and systematization," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2203–2212, Dec. 2011.

[9] G. Ellis and A. Dix, "A taxonomy of clutter reduction for information visualisation," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1216–1223, Nov./Dec. 2007.

[10] P. E. Hoffman and G. G. Grinstein, "A survey of visualizations for high-dimensional data mining," in *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco, CA, USA: Morgan Kaufmann, 2002, pp. 47–82.

[11] D. A. Keim, "Information visualization and visual data mining," *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 1–8, Jan.–Mar. 2002.

[12] M. C. F. De Oliveira and H. Levkowitz, "From visual data exploration to visual data mining: A survey," *IEEE Trans. Vis. Comput. Graph.*, vol. 9, no. 3, pp. 378–394, Jul.–Sep. 2003.

[13] A. Buja, D. Cook, and D. F. Swayne, "Interactive high-dimensional data visualization," *J. Comput. Graphical Statist.*, vol. 5, no. 1, pp. 78–99, 1996.

[14] R. Bürger and H. Hauser, "Visualization of multi-variate scientific data," in *Eurographics State of the Art Reports*. Geneva, Switzerland: Eurographics Assoc., 2007, pp. 117–134.

[15] J. Kehrer and H. Hauser, "Visualization and visual analysis of multifaceted scientific data: A survey," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 3, pp. 495–513, Mar. 2013.

[16] P. C. Wong and R. D. Bergeron, "30 years of multidimensional multivariate visualization," in *Proc. Sci. Vis. Overviews Methodologies Techn.*, 1994, pp. 3–33.

[17] W. W.-Y. Chan, "A survey on multivariate data visualization," *Dep. Comput. Sci. Eng. Hong Kong Univ. Sci. Tech.*, vol. 8, no. 6, pp. 1–29, 2006.

[18] T. Munzner, *Visualization Analysis and Design*. Boca Raton, FL, USA: CRC Press, 2014.

[19] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. San Mateo, CA, USA: Morgan Kaufmann, 1999.

[20] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci, "Visualizing high-dimensional data: Advances in the past decade," in *Proc. Eurographics Conf. Vis. State Art Rep.*, 2015, pp. 127–147.

[21] F. Beck, S. Koch, and D. Weiskopf, "Visual analysis and dissemination of scientific literature collections with survis," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 180–189, Jan. 2016.

[22] G. E. Rosario, E. A. Rundensteiner, D. C. Brown, M. O. Ward, and S. Huang, "Mapping nominal values to numbers for effective visualization," *Inf. Vis.*, vol. 3, no. 2, pp. 80–95, 2004.

[23] M. O. Ward, "XmdvTool: Integrating multiple methods for visualizing multivariate data," in *Proc. IEEE Vis.*, 1994, pp. 326–333.

[24] F. Bendix, R. Kosara, and H. Hauser, "Parallel sets: Visual analysis of categorical data," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 133–140.

[25] C. Weaver, "Conjunctive visual forms," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 929–936, Nov./Dec. 2009.

[26] J.-F. Im, M. McGuffin, and R. Leung, "GPLOM: The generalized plot matrix for visualizing multidimensional multivariate data," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2606–2614, Dec. 2013.

[27] Z. Zhang, K. McDonnell, E. Zadok, and K. Mueller, "Visual correlation analysis of numerical and categorical data on the correlation map," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 2, pp. 289–303, Feb. 2015.

[28] M. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale, "TimeSpan: Using visualization to explore temporal multi-dimensional data of stroke patients," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 409–418, Jan. 2016.

[29] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2005.

[30] D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang, "iPCA: An interactive system for PCA-based visual analytics," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 767–774, 2009.

[31] Y. Koren and L. Carmel, "Visualization of labeled data using linear transformations," in *Proc. IEEE Symp. Inf. Vis.*, 2003, pp. 121–128.

[32] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[33] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[34] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[35] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[36] M. Williams and T. Munzner, "Steerable, progressive multidimensional scaling," in *Proc. IEEE Symp. Inf. Vis.*, 2004, pp. 57–64.

[37] A. Morrison, G. Ross, and M. Chalmers, "A hybrid layout algorithm for sub-quadratic multidimensional scaling," in *Proc. IEEE Symp. Inf. Vis.*, 2002, pp. 152–158.

[38] A. Morrison and M. Chalmers, "Improving hybrid MDS with pivot-based searching," in *Proc. IEEE Symp. Inf. Vis.*, 2003, pp. 11–11.

[39] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[40] V. De Silva and J. B. Tenenbaum, "Sparse multidimensional scaling using landmark points," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2004.

[41] F. Paulovich, C. Silva, and L. Nonato, "Two-phase mapping for projecting massive data sets," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1281–1290, Nov./Dec. 2010.

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

[43] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang, "Dis-function: Learning distance functions interactively," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2012, pp. 83–92.

[44] M. Gleicher, "Explainers: Expert explorations with crafted projections," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2042–2051, Dec. 2013.

[45] J. H. Lee, K. T. McDonnell, A. Zelenyuk, D. Imre, and K. Mueller, "A structure-based distance metric for high-dimensional space exploration with multidimensional scaling," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 3, pp. 351–364, Mar. 2014.

[46] J. A. Lee and M. Verleysen, "Quality assessment of dimensionality reduction: Rank-based criteria," *Neurocomputing*, vol. 72, no. 7, pp. 1431–1443, 2009.

[47] B. Mokbel, W. Lueks, A. Gisbrecht, and B. Hammer, "Visualizing the quality of dimensionality reduction," *Neurocomputing*, vol. 112, pp. 109–123, 2013.

[48] T. Schreck, T. von Landesberger, and S. Bremm, "Techniques for precision-based visual analysis of projected data," *Inf. Vis.*, vol. 9, no. 3, pp. 181–193, 2010.

[49] S. Liu, B. Wang, P.-T. Bremer, and V. Pascucci, "Distortion-guided structure-driven interactive exploration of high-dimensional data," *Comput. Graph. Forum*, vol. 33, no. 3, pp. 101–110, 2014.

[50] C. Seifert, V. Sabol, and W. Kienreich, "Stress maps: Analysing local phenomena in dimensionality reduction based visualisations," in *Proc. IEEE Int. Symp. Visual Analytics Sci. Technol.*, 2010.

[51] J. Stahnke, M. Dörk, B. Müller, and A. Thom, "Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 629–638, Jan. 2016.

[52] C. Turkay, A. Lundervold, A. J. Lundervold, and H. Hauser, "Representative factor generation for the interactive visual analysis of high-dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2621–2630, Dec. 2012.

[53] X. Yuan, D. Ren, Z. Wang, and C. Guo, "Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2625–2633, Dec. 2013.

[54] S. Cheng and K. Mueller, "The data context map: Fusing data and attributes into a unified display," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 121–130, Jan. 2016.

[55] C.-H. Cheng, A. W. Fu, and Y. Zhang, "Entropy-based subspace clustering for mining numerical data," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 1999, pp. 84–93.

[56] J. E. Nam and K. Mueller, "TripAdvisor-ND: A tourism-inspired high-dimensional space exploration framework with overview and detail," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 2, pp. 291–305, Feb. 2013.

[57] A. Tatu, et al., "Subspace search and visualization to make sense of alternative clusterings in high-dimensional data," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2012, pp. 63–72.

[58] C. Baumgartner, C. Plant, K. Railing, H.-P. Kriegel, and P. Kroger, "Subspace selection for clustering high-dimensional data," in *Proc. 4th IEEE Int. Conf. Data Mining*, 2004, pp. 11–18.

[59] B. J. Ferdosi, H. Buddelmeijer, S. Trager, M. H. Wilkinson, and J. B. Roerdink, "Finding and visualizing relevant subspaces for clustering high-dimensional astronomical data using connected morphological operators," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 35–42.

[60] J. Friedman and J. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. C-23, no. 9, pp. 881–890, Sep. 1974.

[61] R. Vidal and E. Elhamifar, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 35, no. 11, pp. 2765–2781, 2013.

[62] D. Lehmann and H. Theisel, "Optimal sets of projections of high-dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 609–618, Jan. 2016.

[63] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 245–250.

[64] A. Anand, L. Wilkinson, and T. N. Dang, "Visual pattern discovery using random projections," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2012, pp. 43–52.

[65] W. Berger, H. Piringer, P. Filzmoser, and E. Gröller, "Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 911–920, 2011.

[66] T. Torsney-Weir, et al., "Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 1892–1901, Dec. 2011.

[67] H. Piringer, W. Berger, and J. Krasser, "HyperMoVal: Interactive visual validation of regression models for real-time simulation," in *Proc. 12th Eurographics / IEEE—VGTC Conf. Vis.*, 2010, pp. 983–992.

[68] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[69] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: MIT Press, 2006.

[70] I. Demir and R. Westermann, "Progressive high-quality response surfaces for visually guided sensitivity analysis," *Comput. Graph. Forum*, vol. 32, no. 3pt1, pp. 21–30, 2013.

[71] C. K. Reddy, S. Pokharkar, and T. K. Ho, "Generating hypotheses of trends in high-dimensional data skeletons," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2008, pp. 139–146.

[72] T. Hastie and W. Stuetzle, "Principal curves," *J. Amer. Statistical Assoc.*, vol. 84, no. 406, pp. 502–516, 1989.

[73] A. J. Zomorodian, *Topology for Computing (Cambridge Monographs on Applied and Computational Mathematics)*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[74] S. Biasotti, et al., "Describing shapes by geometrical-topological properties of real functions," *ACM Comput. Surveys*, vol. 40, no. 4, pp. 12:1–12:87, 2008.

[75] H. Edelsbrunner and J. Harer, "Persistent homology—A survey," *Contemporary Math.*, vol. 453, pp. 257–282, 2008.

[76] H. Edelsbrunner and J. Harer, *Computational Topology—An Introduction*. Providence, RI, USA: Amer. Math. Soc., 2010.

[77] G. Carlsson, "Topology and data," *Bulletin Amer. Math. Soc.*, vol. 46, no. 2, pp. 255–308, 2009.

[78] R. Ghrist, "Barcodes: The persistent topology of data," *Bulletin Amer. Math. Soc.*, vol. 45, pp. 61–75, 2008.

[79] C. Heine, et al., "A survey of topology-based methods in visualization," *Comput. Graph. Forum*, vol. 35, pp. 643–667, 2016.

[80] R. Fuchs and H. Hauser, "Visualization of multi-variate scientific data," *Comput. Graph. Forum*, vol. 28, no. 6, pp. 1670–1690, 2009.

[81] R. S. Laramee, G. Chen, M. Jankun-Kelly, E. Zhang, and D. Thompson, *Bringing Topology-Based Flow Visualization to the Application Domain*. Berlin, Germany: Springer, 2009, pp. 161–176.

[82] R. S. Laramee, H. Hauser, L. Zhao, and F. H. Post, *Topology-Based Flow Visualization, The State of the Art*. Berlin, Germany: Springer, 2007, pp. 1–19.

[83] T. Peacock and J. Dabiri, "Introduction to focus issue: Lagrangian coherent structures," *Chaos*, vol. 20, no. 017501, 2010.

[84] A. Pobitzer, et al., "The state of the art in topology-based visualization of unsteady flow," *Comput. Graph. Forum*, vol. 30, no. 6, pp. 1789–1811, 2011.

[85] F. H. Post, B. Vrolijk, H. Hauser, R. S. Laramee, and H. Doleisch, "The state of the art in flow visualisation: Feature extraction and tracking," *Comput. Graph. Forum*, vol. 22, no. 4, pp. 775–792, 2003.

[86] W. Wang, W. Wang, and S. Li, "From numerics to combinatorics: A survey of topological methods for vector field visualization," *J. Vis.*, vol. 19, no. 4, pp. 727–752, 2016.

[87] M. Hlawatsch, F. Sadlo, and D. Weiskopf, "Hierarchical line integration," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 8, pp. 1148–1163, Aug. 2011.

[88] G. Reeb, "Sur les points singuliers d'une forme de pfaff completement intergrable ou d'une fonction numerique [on the singular points of a complete integral pfaff form or of a numerical function]," *Comptes Rendus Acad. Sci. Paris*, vol. 222, pp. 847–849, 1946.

[89] S. Smale, "On gradient dynamical systems," *Ann. Math.*, vol. 74, pp. 199–206, 1961.

[90] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci, "Morse-Smale complexes for piece-wise linear 3-manifolds," in *Proc. 19th Annu. Symp. Comput. Geometry*, 2003, pp. 361–370.

[91] H. Edelsbrunner, J. Harer, and A. J. Zomorodian, "Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds," *Discr. Comput. Geometry*, vol. 30, pp. 87–107, 2003.

[92] L. Čomić and L. D. Floriani, "Dimension-independent simplification and refinement of Morse complexes," *Graphical Models*, vol. 73, no. 5, pp. 261–285, 2011.

[93] D. Günther, J. Reininghaus, H.-P. Seidel, and T. Weinkauf, "Notes on the simplification of the Morse-Smale complex," in *Proc. Topological Methods Data Anal. Vis.*, Mar. 2013, pp. 135–150.

[94] F. Iuricich, U. Fugacci, and L. D. Floriani, "Topologically-consistent simplification of discrete Morse complex," *Comput. Graph.*, vol. 51, pp. 157–166, 2015.

[95] D. Maljovec, et al., "Analyzing simulation-based PRA data through traditional and topological clustering: A BWR station blackout case study," *Rel. Eng. Syst. Safety*, vol. 145, pp. 262–276, 2016.

[96] D. Maljovec, et al., "Exploration of high-dimensional scalar function for nuclear reactor safety analysis and visualization," in *Proc. Int. Conf. Math. Comput. Methods Appl. Nuclear Sci. Eng.*, 2013, pp. 712–723.

[97] P.-T. Bremer, et al., "ND2AV: N-dimensional data analysis and visualization analysis for the national ignition campaign," *Comput. Vis. Sci.*, vol. 17, pp. 1–18, 2014.

[98] C. Correa and P. Lindstrom, "Towards robust topology of sparsely sampled data," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 1852–1861, Dec. 2011.

[99] C. D. Correaa, P. Lindstrom, and P.-T. Bremer, "Topological spines: A structure-preserving visual representation of scalar fields," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 1842–1851, Dec. 2011.

[100] V. Narayanan, D. Thomas, and V. Natarajan, "Distance between extremum graphs," in *Proc. IEEE Pacific Vis. Symp.*, Apr. 2015, pp. 263–270.

[101] M. Nicolau, A. J. Levine, and G. Carlsson, "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival," *Proc. Nat. Academy Sci. USA*, vol. 108, no. 17, pp. 7265–7270, 2011.

[102] G. Singh, F. Memoli, and G. Carlsson, "Topological methods for the analysis of high dimensional data sets and 3D object recognition," in *Proc. Symp. Point Based Graph.*, 2007, pp. 91–100.

[103] T. K. Dey and Y. Wang, "Reeb graphs: Approximation and persistence," *Discr. Comput. Geometry*, vol. 49, no. 1, pp. 46–73, 2013.

[104] S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno, "Reeb graphs for shape analysis and applications," *Theoretical Comput. Sci.*, vol. 392, no. 1, pp. 5–22, 2008.

[105] G. Sarikonda, et al., "CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes," *J. Autoimmunity*, vol. 50, pp. 77–82, 2014.

[106] E. Munch and B. Wang, "Convergence between categorical representations of reeb space and mapper," in *Proc. Int. Symp. Comput. Geometry*, 2016.

[107] H. Carr, Z. Geng, J. Tierny, A. Chattopadhyay, and A. Knoll, "Fiber surfaces: Generalizing isosurfaces to bivariate data," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 241–250, 2015.

[108] P. Klacansky, J. Tierny, H. Carr, and Z. Geng, "Fast and exact fiber surfaces for tetrahedral meshes," *IEEE Trans. Vis. Comput. Graph.*, vol. PP, no. 99, p. 1, 2016.

[109] J. Tierny and H. Carr, "Jacobi fiber surfaces for bivariate reeb space computation," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 960–969, Jan. 2016.

[110] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discr. Comput. Geometry*, vol. 37, no. 1, pp. 103–120, 2007.

[111] K. Beketayev, D. Yeliussizov, D. Morozov, G. Weber, and B. Hamann, "Measuring the distance between merge trees," in *Topological Methods in Data Analysis and Visualization III*, P.-T. Bremer, I. Hotz, V. Pascucci, and R. Peikert, Eds. Berlin, Germany: Springer, 2014, pp. 151–165.

[112] H. Carr, J. Snoeyink, and U. Axen, "Computing contour trees in all dimensions," *Comput. Geometry*, vol. 24, no. 2, pp. 75–94, 2003.

[113] Y.-J. Chiang, T. Lenz, X. Lu, and G. Rote, "Simple and optimal output-sensitive construction of contour trees using monotone paths," *Comput. Geometry*, vol. 30, no. 2, pp. 165–195, 2005.

[114] B. Raichel and C. Seshadhri, "A mountaintop view requires minimal sorting: A faster contour tree algorithm," *Preprint*, vol. arXiv:1411.2689v4 [cs.CG], 2014.

[115] P. Oesterling, C. Heine, G. Weber, D. Morozov, and G. Scheuermann, "Computing and visualizing time-varying merge trees for high-dimensional data," in *Proc. Workshop on Topology-Based Methods Vis.*, May 2015.

[116] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas, "Robust on-line computation of Reeb graphs: Simplicity and speed," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 58.

[117] H. Carr, C. Sewell, L.-T. Lo, and J. Ahrens, "Hybrid data-parallel contour tree computation," in *Computer Graphics and Visual Computing*, C. Turkay and T. R. Wan, Eds. Geneva, Switzerland: Eurographics Assoc., 2016.

[118] H. Carr, G. Weber, C. Sewell, and J. Ahrens, "Parallel peak pruning for scalable SMP contour tree computation," in *Proc. IEEE Symp. Large Data Anal. Vis.*, 2016.

[119] S. Maadasamy, H. Doraiswamy, and V. Natarajan, "A hybrid parallel algorithm for computing and tracking level set topology," in *Proc. IEEE 19th Int. Conf. High Perform. Comput.*, 2012, pp. 1–10.

[120] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny, "Contour forests: Fast multi-threaded augmented contour trees," in *Proc. IEEE Symp. Large Data Anal. Vis.*, 2016.

[121] A. Mascarenhas and J. Snoeyink, *Isocontour Based Visualization of Time-Varying Scalar Fields*. Berlin, Germany: Springer, 2009, pp. 41–68.

[122] H. Edelsbrunner and J. Harer, "Jacobi sets of multiple morse functions," *Found. Comput. Math.*, vol. 312, pp. 37–57, 2002.

[123] H. Carr and D. Duke, "Joint contour nets," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 8, pp. 1100–1113, Aug. 2014.

[124] D. Duke, H. Carr, A. Knoll, N. Schunck, H. A. Nam, and A. Staszczak, "Visualizing nuclear scission through a multifield extension of topological analysis," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2033–2040, Dec. 2012.

[125] D. J. Duke and F. Hosseini, "Skeletons for distributed topological computation," in *Proc. 4th ACM SIGPLAN Workshop Functional High-Perform. Comput.*, 2015, pp. 35–44.

[126] Z. Geng, D. Duke, and H. Carr, "Topological analysis of atlantic ocean data using joint contour net," in *Proc. Workshop Topology-Based Methods Vis.*, May 2015.

[127] A. Chattopadhyay, H. Carr, D. Duke, and Z. Geng, "Extracting Jacobi structures in Reeb spaces," in *EuroVis—Short Papers*, N. Elmqvist, M. Hlawitschka, and J. Kennedy, Eds. Geneva, Switzerland: Eurographics Assoc., 2014.

[128] A. Chattopadhyay, H. Carr, D. Duke, Z. Geng, and O. Saeki, "Multivariate topology simplification," *Comput. Geometry*, vol. 58, pp. 1–24, 2016.

[129] L. Huettenberger, C. Heine, H. Carr, G. Scheuermann, and C. Garth, "Towards multifield scalar topology based on pareto optimality," *Comput. Graph. Forum*, vol. 32, no. 3pt3, pp. 341–350, 2013.

[130] P. Stadler and C. Flamm, "Barrier trees on poset-valued landscapes," *Genetic Program. Evolvable Mach.*, vol. 4, no. 1, pp. 7–20, 2003.

[131] L. Huettenberger, C. Heine, and C. Garth, "Decomposition and simplification of multivariate data using pareto sets," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2684–2693, Dec. 2014.

[132] L. Huettenberger, C. Heine, and C. Garth, "A comparison of joint contour nets and pareto sets," in *Proc. Workshop Topology-Based Methods Vis.*, May 2015.

[133] A. B. Lee, K. S. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," *Int. J. Comput. Vis.*, vol. 54, no. 1–3, pp. 83–103, 2003.

[134] B. Wang, B. Summa, V. Pascucci, and M. Vejdemo-Johansson, "Branching and circular features in high dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 1902–1911, Dec. 2011.

[135] V. de Silva, D. Morozov, and M. Vejdemo-Johansson, "Persistent cohomology and circular coordinates," in *Proc. 25th Annu. Symp. Comput. Geometry*, 2009, pp. 227–236.

[136] B. Rieck, H. Mara, and H. Leitte, "Multivariate data analysis using persistence-based filtering and topological signatures," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2382–2391, Dec. 2012.

[137] B. Rieck and H. Leitte, "Structural analysis of multivariate point clouds using simplicial chains," *Comput. Graph. Forum*, vol. 33, no. 8, pp. 28–37, 2014.

[138] B. Rieck and H. Leitte, "Persistent homology for the evaluation of dimensionality reduction schemes," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 431–440, 2015.

[139] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 77–102, 2015.

[140] L. Wilkinson, A. Anand, and R. Grossman, "Graph-theoretic scagnostics," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 157–164.

[141] T. N. Dang, A. Anand, and L. Wilkinson, "TimeSeer: Scagnostics for high-dimensional time series," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 3, pp. 470–483, Mar. 2013.

[142] D. Guo, "Coordinating computational and visual approaches for interactive feature selection and multivariate clustering," *Inf. Vis.*, vol. 2, no. 4, pp. 232–246, 2003.

[143] J. Seo and B. Shneiderman, "A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections," in *Proc. IEEE Symp. Inf. Vis.*, 2004, pp. 65–72.

[144] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan, "Selecting good views of high-dimensional data using class consistency," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 831–838, 2009.

[145] A. Tatu, et al., "Combining automated analysis and visualization techniques for effective exploration of high-dimensional data," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2009, pp. 59–66.

[146] D. J. Lehmann, G. Albuquerque, M. Eisemann, M. Magnor, and H. Theisel, "Selecting coherent and relevant plots in large scatterplot matrices," *Comput. Graph. Forum*, vol. 31, no. 6, pp. 1895–1908, 2012.

[147] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multi-dimensional geometry," in *Human-Machine Interactive Systems*. Berlin, Germany: Springer, 1991, pp. 199–233.

[148] C. Hurley and R. Oldford, "Pairwise display of high-dimensional information via eulerian tours and hamiltonian decompositions," *J. Comput. Graph. Statist.*, vol. 19, no. 4, pp. 861–886, 2010.

[149] B. J. Ferdosi and J. B. Roerdink, "Visualizing high-dimensional structures by dimension ordering and filtering using subspace analysis," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 1121–1130, 2011.

[150] S. Johansson and J. Johansson, "Interactive dimensionality reduction through user-defined combinations of quality metrics," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 993–1000, Nov./Dec. 2009.

[151] M. Novotny and H. Hauser, "Outlier-preserving focus+context visualization in parallel coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 893–900, Sep./Oct. 2006.

[152] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen, "Visual clustering in parallel coordinates," *Comput. Graph. Forum*, vol. 27, no. 3, pp. 1047–1054, 2008.

[153] R. Rosenbaum, J. Zhi, and B. Hamann, "Progressive parallel coordinates," in *Proc. IEEE Pacific Vis. Symp.*, 2012, pp. 25–32.

[154] T. N. Dang, L. Wilkinson, and A. Anand, "Stacking graphic elements to avoid over-plotting," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1044–1052, Nov./Dec. 2010.

[155] W. Peng, M. O. Ward, and E. A. Rundensteiner, "Clutter reduction in multi-dimensional data visualization using dimension reordering," in *Proc. IEEE Symp. Inf. Vis.*, 2004, pp. 89–96.

[156] E. Kandogan, "Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions," in *Proc. IEEE Inf. Vis. Symp. Late Breaking Hot Topics*, 2000, pp. 9–12.

[157] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley, "DNA visual and analytic data mining," in *Proc. IEEE Vis.*, 1997, pp. 437–441.

[158] D. J. Lehmann and H. Theisel, "Orthographic star coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2615–2624, Dec. 2013.

[159] G. Albuquerque, M. Eisemann, D. J. Lehmann, H. Theisel, and M. Magnor, "Improving the visual analysis of high-dimensional datasets using quality measures," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 19–26.

[160] N. Elmqvist, J. Stasko, and P. Tsigas, "Datameadow: A visual canvas for analysis of large-scale multivariate data," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2007, pp. 187–194.

[161] S. Jayaraman and C. North, "A radial focus+context visualization for multi-dimensional functions," in *Proc. IEEE Vis.*, 2002, pp. 443–450.

[162] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu, "Scattering points in parallel coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1001–1008, Nov./Dec. 2009.

[163] J. Claessen and J. van Wijk, "Flexible linked axes for multivariate data visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2310–2316, Dec. 2011.

[164] E. Fanea, S. Carpendale, and T. Isenberg, "An interactive 3D integration of parallel coordinates and star glyphs," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 149–156.

[165] Z. Geng, Z. Peng, R. Laramee, J. Roberts, and R. Walker, "Angular histograms: Frequency-based visualizations for large, high dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2572–2580, Dec. 2011.

[166] M. Caat, N. Maurits, and J. Roerdink, "Design and evaluation of tiled parallel coordinates visualization of multichannel EEG data," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 1, pp. 70–79, Jan./Feb. 2007.

[167] H. Chernoff, "The use of faces to represent points in k-dimensional space graphically," *J. Amer. Statistical Assoc.*, vol. 68, no. 342, pp. 361–368, 1973.

[168] C. Correa, Y.-H. Chan, and K.-L. Ma, "A framework for uncertainty-aware visual analytics," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2009, pp. 51–58.

[169] Y.-H. Chan, C. Correa, and K.-L. Ma, "Flow-based scatterplots for sensitivity analysis," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 43–50.

[170] Z. Guo, M. Ward, E. Rundensteiner, and C. Ruiz, "Pointwise local pattern exploration for sensitivity analysis," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2011, pp. 131–140.

[171] Y.-H. Chan, C. Correa, and K.-L. Ma, "The generalized sensitivity scatterplot," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 10, pp. 1768–1781, Oct. 2013.

[172] N. Cao, D. Gotz, J. Sun, and H. Qu, "DICON: Interactive visual analysis of multidimensional clusters," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2581–2590, Dec. 2011.

[173] D. J. Lehmann, F. Kemmler, T. Zhyhalava, M. Kirschke, and H. Theisel, "Visualnostics: Visual guidance pictograms for analyzing projections of high-dimensional data," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 291–300, 2015.

[174] M. O. Ward, "Multivariate data glyphs: Principles and practice," in *Handbook of Data Visualization*. Berlin, Germany: Springer, 2008, pp. 179–198.

[175] D. Keim and H.-P. Kriegel, "VisDB: Database exploration using multidimensional visualization," *IEEE Comput. Graph. Appl.*, vol. 14, no. 5, pp. 40–49, Sep. 1994.

[176] M. Ankerst, D. A. Keim, and H. peter Kriegel, "Circle segments: A technique for visually exploring large multidimensional data sets," in *Proc. IEEE Vis. Hot Topic Session*, 1996.

[177] D. Keim, M. Hao, J. Ladisch, M. Hsu, and U. Dayal, "Pixel bar charts: A new technique for visualizing large multi-attribute data sets without aggregation," in *Proc. IEEE Symp. Inf. Vis.*, 2001, pp. 113–120.

[178] M. Wattenberg, "A note on space-filling visualizations and space-filling curves," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 181–186.

[179] J. Yang, A. Patro, H. Shiping, N. Mehta, M. Ward, and E. Rundensteiner, "Value and relation display for interactive exploration of high dimensional datasets," in *Proc. IEEE Symp. Inf. Vis.*, 2004, pp. 73–80.

[180] D. Keim, H.-P. Kriegel, and M. Ankerst, "Recursive pattern: A technique for visualizing very large amounts of data," in *Proc. IEEE Vis.*, 1995, pp. 279–286.

[181] J. Yang, D. Hubball, M. O. Ward, E. A. Rundensteiner, and W. Ribarsky, "Value and relation display: Interactive visual exploration of large data sets with hundreds of dimensions," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 3, pp. 494–507, May 2007.

[182] J. Wang, W. Peng, M. O. Ward, and E. A. Rundensteiner, "Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets," in *Proc. IEEE Symp. Inf. Vis.*, 2003, pp. 105–112.

[183] J. Yang, M. O. Ward, and E. A. Rundensteiner, "Interring: An interactive tool for visually navigating and manipulating hierarchical structures," in *Proc. IEEE Symp. Inf. Vis.*, 2002, pp. 77–84.

[184] C. Heine, D. Schneider, H. Carr, and G. Scheuermann, "Drawing contour trees in the plane," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 11, pp. 1599–1611, Nov. 2011.

[185] V. Pascucci, K. Cole-McLaughlin, and G. Scorzelli, "The toporrery: Computation and presentation of multi-resolution topology," in *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*. Berlin, Germany: Springer, 2009, pp. 19–40.

[186] G. H. Weber, P.-T. Bremer, and V. Pascucci, "Topological cacti: Visualizing contour-based statistics," in *Topological Methods in Data Analysis and Visualization II*. Berlin, Germany: Springer, 2012, pp. 63–76.

[187] K. Beketayev, D. Morozov, G. H. Weber, A. Abzhanov, and B. Hamann, "Geometry-preserving topological landscapes," in *Proc. Workshop SIGGRAPH Asia*, 2012, pp. 155–160.

[188] D. Demir, K. Beketayev, G. H. Weber, P.-T. Bremer, V. Pascucci, and B. Hamann, "Topology exploration with hierarchical landscapes," in *Proc. Workshop SIGGRAPH Asia*, 2012, pp. 147–154.

[189] W. Harvey and Y. Wang, "Topological landscape ensembles for visualization of scalar-valued functions," *Comput. Graph. Forum*, vol. 29, no. 3, pp. 993–1002, 2010.

[190] P. Oesterling, C. Heine, H. Jänicke, and G. Scheuermann, "Visual analysis of high dimensional point clouds using topological landscape," in *Proc. IEEE Pacific Vis. Symp.*, 2010, pp. 113–120.

[191] P. Oesterling, C. Heine, H. Janicke, G. Scheuermann, and G. Heyer, "Visualization of high-dimensional point clouds using their density distribution's topology," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 11, pp. 1547–1559, Nov. 2011.

[192] G. Weber, P.-T. Bremer, and V. Pascucci, "Topological landscapes: A terrain metaphor for scientific data," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1416–1423, Nov./Dec. 2007.

[193] P. Oesterling, C. Heine, G. Weber, and G. Scheuermann, "Visualizing nD point clouds as topological landscape profiles to guide local data analysis," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 3, pp. 514–526, Mar. 2013.

[194] J. Klemelä, "Visualization of multivariate density estimates with level set trees," *J. Comput. Graphical Statist.*, vol. 13, no. 3, pp. 599–620, 2004.

[195] K. Beketayev, G. Weber, M. Haranczyk, P.-T. Bremer, M. Hlawitschka, and B. Hamann, "Topology-based visualization of transformation pathways in complex chemical data," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 663–672, 2011.

[196] A. Gyulassy, V. Natarajan, V. Pascucci, and B. Hamann, "Efficient computation of Morse-Smale complexes for three-dimensional scalar functions," *IEEE Trans. Comput. Graph. Vis.*, vol. 13, no. 6, pp. 1440–1447, Nov./Dec. 2007.

[197] V. Pascucci, K. Cole-McLaughlin, and G. Scorzelli, "Multi-resolution computation and presentation of contour trees," Lawrence Livermore Nat. Laboratory, Livermore, CA, USA, Tech. Rep. UCRL-PROC-208680, 2005.

[198] Y.-H. Fua, M. Ward, and E. Rundensteiner, "Structure-based brushes: A mechanism for navigating hierarchically organized data and information spaces," *IEEE Trans. Vis. Comput. Graph.*, vol. 6, no. 2, pp. 150–159, Apr.–Jun. 2000.

[199] B. Shneiderman, "Tree visualization with tree-maps: 2-d space-filling approach," *ACM Trans. Graph.*, vol. 11, no. 1, pp. 92–99, 1992.

[200] D. Engel, R. Rosenbaum, B. Hamann, and H. Hagen, "Structural decomposition trees," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 921–930, 2011.

[201] M. Kreuseler and H. Schumann, "A flexible approach for visual data mining," *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 39–51, Jan.–Mar. 2002.

[202] J. Heer and G. G. Robertson, "Animated transitions in statistical data graphics," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1240–1247, Nov./Dec. 2007.

[203] D. F. Swayne, D. T. Lang, A. Buja, and D. Cook, "GGobi: Evolving from XGobi into an extensible framework for interactive data visualization," *Comput. Statist. Data Anal.*, vol. 43, no. 4, pp. 423–444, 2003.

[204] N. Elmqvist, P. Dragicevic, and J.-D. Fekete, "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1539–1148, Nov./Dec. 2008.

[205] A. Waddell and R. W. Oldford, "RnavGraph: A visualization tool for navigating through high-dimensional data," in *Proc. 58th World Statistical Congr.*, 2011, pp. 1852–1860.

[206] M. Sedlmair, T. Munzner, and M. Tory, "Empirical guidance on scatterplot and dimension reduction technique choices," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2634–2643, Dec. 2013.

[207] R. Etemadpour, R. Motta, J. de Souza Paiva, R. Minghim, M. Ferreira de Oliveira, and L. Linsen, "Perception-based evaluation of projection methods for multidimensional data visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 1, pp. 81–94, Jan. 2015.

[208] D. Holten and J. J. Van Wijk, "Evaluation of cluster identification performance for different PCP variants," *Comput. Graph. Forum*, vol. 29, no. 3, pp. 793–802, 2010.

[209] M. Rubio-Sanchez, L. Raya, F. Diaz, and A. Sanchez, "A comparative study between RadViz and star coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 619–628, Jan. 2016.

[210] Y. Albo, J. Lanir, P. Bak, and S. Rafaeli, "Off the radar: Comparative evaluation of radial visualization solutions for composite indicators," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 569–578, Jan. 2016.

[211] H. Sanftmann and D. Weiskopf, "Illuminated 3D scatterplots," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 751–758, 2009.

[212] K. T. McDonnell and K. Mueller, "Illustrative parallel coordinates," *Comput. Graph. Forum*, vol. 27, no. 3, pp. 1031–1038, 2008.

[213] J. Johansson, P. Ljung, M. Jern, and M. Cooper, "Revealing structure within clustered parallel coordinates displays," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 125–132.

[214] R. Rao and S. K. Card, "The table lens: Merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information," in *Proc. ACM SIGCHI Conf. Human Factors Comput. Syst.*, 1994, pp. 318–322.

[215] C. Hurter, A. Telea, and O. Ersoy, "MoleView: An attribute and structure-based semantic lens for large element-based plots," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2600–2609, Dec. 2011.

[216] Y. K. Leung and M. D. Apperley, "A review and taxonomy of distortion-oriented presentation techniques," *ACM Trans. Comput.-Human Interaction*, vol. 1, no. 2, pp. 126–160, 1994.

[217] S. Bachthaler and D. Weiskopf, "Continuous scatterplots," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1428–1435, Nov./Dec. 2008.

[218] S. Bachthaler and D. Weiskopf, "Efficient and adaptive rendering of 2-D continuous scatterplots," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 743–750, 2009.

[219] J. Heinrich and D. Weiskopf, "Continuous parallel coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1531–1538, Nov./Dec. 2009.

[220] D. Lehmann and H. Theisel, "Features in continuous parallel coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 1912–1921, Dec. 2011.

[221] A. Mayorga and M. Gleicher, "Splatterplots: Overcoming overdraw in scatter plots," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 9, pp. 1526–1538, Sep. 2013.

[222] H. Zhou, W. Cui, H. Qu, Y. Wu, X. Yuan, and W. Zhuo, "Splatting the lines in parallel coordinates," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 759–766, 2009.

[223] L. Kuhne, J. Giesen, Z. Zhang, S. Ha, and K. Mueller, "A data-driven approach to hue-preserving color-blending," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2122–2129, Dec. 2012.

[224] H. Hagh-Shenas, S. Kim, V. Interrante, and C. Healey, "Weaving versus blending: A quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1270–1277, Nov./Dec. 2007.

[225] A. Artero, M. de Oliveira, and H. Levkowitz, "Uncovering clusters in crowded parallel coordinates visualizations," in *Proc. IEEE Symp. Inf. Vis.*, 2004, pp. 81–88.

[226] J. Johansson and M. Cooper, "A screen space quality method for data abstraction," *Comput. Graph. Forum*, vol. 27, no. 3, pp. 1039–1046, 2008.

[227] A. Dasgupta and R. Kosara, "Pargnostics: Screen-space metrics for parallel coordinates," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1017–1026, Nov./Dec. 2010.

[228] J. Schneidewind, M. Sips, and D. A. Keim, "Pixnostics: Towards measuring the value of visualization," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2006, pp. 199–206.

[229] T. Muhlbacher, H. Piringer, S. Gratzl, M. Sedlmair, and M. Streit, "Opening the black box: Strategies for increased user involvement in existing algorithm implementations," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1643–1652, Dec. 2014.

[230] C. Turkay, F. Jeanquartier, A. Holzinger, and H. Hauser, "On computationally-enhanced visual analysis of heterogeneous data and its application in biomedical informatics," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Berlin, Germany: Springer, 2014, pp. 117–140.

[231] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit, "LineUp: Visual analysis of multi-attribute rankings," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2277–2286, Dec. 2013.

[232] S. Gratzl, N. Gehlenborg, A. Lex, H. Pfister, and M. Streit, "Domino: Extracting, comparing, and manipulating subsets across multiple tabular datasets," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2023–2032, Dec. 2014.

[233] D. A. Keim and H.-P. Kriegel, "Visualization techniques for mining large databases: A comparison," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 923–938, Dec. 1996.

[234] X. Hu, L. Bradel, D. Maiti, L. House, and C. North, "Semantics of directly manipulating spatializations," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2052–2059, Dec. 2013.

[235] H. Kim, J. Choo, H. Park, and A. Endert, "InterAxis: Steering scatterplot axes via observation-level interaction," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 131–140, Jan. 2016.

[236] C. R. Johnson, "Top scientific visualization research problems," *IEEE Comput. Graph. Appl.*, vol. 24, no. 4, pp. 13–17, Jul.–Aug. 2004.

[237] K. Brodlie, R. A. Osorio, and A. Lopes, "A review of uncertainty in data visualization," in *Expanding the Frontiers of Visual Analytics and Visualization*, J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Eds. London, U.K.: Springer, 2012, pp. 81–109.

[238] H. Griethe and H. Schumann, "The visualization of uncertain data: Methods and problems," in *Proc. SimVis*, 2006, pp. 143–156.

[239] A. Pang, C. Wittenbrink, and S. Lodha, "Approaches to uncertainty visualization," *Visual Comput.*, vol. 13, no. 8, pp. 370–390, 1997.

[240] K. Potter, P. Rosen, and C. R. Johnson, "From quantification to visualization: A taxonomy of uncertainty visualization approaches," *IFIP Advances Inf. Commun. Technol. Series*, vol. 377, pp. 226–249, 2012.

[241] A. Cedilnik and P. Rheingans, "Procedural annotation of uncertain information," in *Proc. 11th IEEE Vis. Conf.*, 2000, pp. 77–84.

[242] D. Feng, L. Kwock, Y. Lee, and R. Taylor, "Matching visual saliency to confidence in plots of uncertain data," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 980–989, Nov. 2010.

[243] G. Grigoryan and P. Rheingans, "Probabilistic surfaces: Point based primitives to show surface uncertainty," in *Proc. IEEE Vis.*, 2002, pp. 147–153.

[244] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha, "Glyphs for visualizing uncertainty in vector fields," *IEEE Trans. Vis. Comput. Graph.*, vol. 2, no. 3, pp. 266–279, Sep. 1996.

[245] A. Coninx, G.-P. Bonneau, J. Droulez, and G. Thibault, "Visualization of uncertain scalar data fields using color scales and perceptually adapted noise," in *Proc. ACM SIGGRAPH Symp. Appl. Perception Graph. Vis.*, 2011, pp. 59–66.

[246] S. Djurcilov, K. Kim, P. Lermusiaux, and A. Pang, "Visualizing scalar volumetric data with uncertainty," *Comput. Graph.*, vol. 26, pp. 239–248, 2002.

[247] H. Chen, et al., "Uncertainty-aware multidimensional ensemble data visualization and exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 9, pp. 1072–1086, Sep. 2015.

[248] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, V. Pascucci, and C. Johnson, "Ensemblevis: A flexible approach for the statistical visualization of ensemble data," in *Proc. IEEE Workshop Knowl. Discovery Climate Data: Prediction Extremes Impacts*, 2009.

[249] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead, "Noodles: A tool for visualization of numerical weather model ensemble uncertainty," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1421–1430, Nov./Dec. 2010.

[250] Y. Wang and K.-L. Ma, "Revealing the fog-of-war: A visualization-directed, uncertainty-aware approach for exploring high-dimensional data," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2015, pp. 629–638.

[251] X. Zaixian, H. Shiping, M. Ward, and E. Rundensteiner, "Exploratory visualization of multivariate data with variable quality," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2006, pp. 183–190.

[252] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo, "VisTrails: Visualization meets data management," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2006, pp. 745–747.

[253] S. Berchtold, D. Keim, and H. Kriegel, "The X-tree: An index structure for high-dimensional data," in *Readings in Multimedia Computing and Networking*. San Francisco, CA, USA: Morgan Kaufmann, 2001, pp. 451–462.

[254] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA, USA: Addison-Wesley, 1977.

[255] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, *Visual Analytics: Definition, Process, and Challenges*. Berlin, Germany: Springer, 2008.

[256] D. A. Keim, F. Mansmann, and J. Thomas, "Visual analytics: How much visualization and how much analytics?" *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 2, pp. 5–8, 2010.

[257] C. Stolte, D. Tang, and P. Hanrahan, "Polaris: A system for query, analysis, and visualization of multidimensional relational databases," *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 52–65, Jan.–Mar. 2002.

[258] C. Stolte, D. Tang, and P. Hanrahan, "Query, analysis, and visualization of hierarchically structured data using polaris," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2002, pp. 112–122.

[259] C. Stolte, D. Tang, and P. Hanrahan, "Multiscale visualization using data cubes," *IEEE Trans. Vis. Comput. Graph.*, vol. 9, no. 2, pp. 176–187, Apr.–Jun. 2003.

[260] M. Hao, U. Dayal, D. Keim, D. Morent, and J. Schneidewind, "Intelligent visual analytics queries," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2007, pp. 91–98.

[261] S. Liu, B. Wang, J. J. Thiagarajan, P.-T. Bremer, and V. Pascucci, "Visual exploration of high-dimensional data through subspace analysis and dynamic Projections," *Comput. Graph. Forum*, vol. 34, pp. 271–280, 2015.

[262] F.-Y. Tzeng and K.-L. Ma, "Opening the black box-data driven visualization of neural networks," in *Proc. IEEE Vis.*, 2005, pp. 383–390.

[263] S. T. Teoh and K.-L. Ma, "PaintingClass: Interactive construction, visualization and exploration of decision trees," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 667–672.

[264] S. van den Elzen and J. van Wijk, "BaobabView: Interactive construction and analysis of decision trees," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2011, pp. 151–160.

[265] P. Rheingans and M. desJardins, "Visualizing high-dimensional predictive model quality," in *Proc. IEEE Vis.*, 2000, pp. 493–496.

[266] M. Migut and M. Worring, "Visual exploration of classification models for risk assessment," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 11–18.

[267] S. Garg, I. Ramakrishnan, and K. Mueller, "A visual analytics approach to model learning," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 67–74.

[268] J. Seo and B. Shneiderman, "Knowledge discovery in high-dimensional data: Case studies and a user survey for the rank-by-feature framework," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 3, pp. 311–322, May/Jun. 2006.

Shusen Liu received the bachelor's degree in biomedical engineering and computer science from Huazhong University of Science and Technology, China, in 2009, where he worked in Wuhan National Laboratory for Optoelectronic on GPU accelerated biophotonics applications. He is currently working toward the PhD degree at University of Utah. His research interests lie primarily in high-dimensional data visualization and multivariate volume visualization.

Dan Maljovec received the BS degree in computer science from Gannon University, in 2009. He is working toward the PhD degree in the School of Computing, University of Utah. He has been a research assistant with the University of Utah's Scientific Computing and Imaging Institute since 2012. His research focuses on analysis and visualization of high-dimensional scientific data and intuitive visualization.

Bei Wang received the PhD degree in computer science from Duke University, in 2010. She is currently an assistant professor in the Scientific Computing and Imaging Institute, University of Utah. Her research expertise lies in the theoretical, algorithmic, and application aspects of data analysis and data visualization, with a focus on topological techniques. She is a member of the IEEE Computer Society and the ACM.

Peer-Timo Bremer received the Diploma degree in mathematics and computer science from the Leipniz University, Hannover, Germany, in 2000, and the PhD degree in computer science from the University of California, Davis, in 2004. He is a project leader in the Center for Applied Scientific Computing, Lawrence Livermore National Laboratory. His research interests include large scale data analysis, performance analysis, and visualization. He is a member of the IEEE Computer Society and the ACM.

Valerio Pascucci received the EE Laurea (master's) degree from the University "La Sapienza", Rome, Italy, and the PhD degree in computer science from Purdue University, in May 2000. He is the founding director of the Center for Extreme Data Management Analysis and Visualization (CEDMAV), University of Utah. He is also a faculty of the Scientific Computing and Imaging Institute, a professor of the School of Computing, University of Utah, and a DOE laboratory fellow of the Pacific Northwest National Laboratory. His recent research interest is in developing new methods for massive data analysis and visualization. He is a member of the IEEE.