Eurographics Conference on Visualization (EuroVis) (2015)
R. Borgo, F. Ganovelli, and I. Viola (Editors)

Visualizing High-Dimensional Data: Advances in the Past Decade

S. Liu[1], D. Maljovec[1], B. Wang[1], P.-T. Bremer[2] and V. Pascucci[1]

[1]Scientific Computing and Imaging Institute, University of Utah
[2]Lawrence Livermore National Laboratory

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 over the past 15 years. We aim at providing actionable guidance for data practitioners to navigate through a modular view of the recent advances, allowing the creation of new visualizations along the enriched information visualization pipeline and identifying future opportunities for visualization research.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1 Introduction

With the ever-increasing amount of available computing resources, our ability to collect and generate a wide variety of large, complex, high-dimensional datasets continues to grow. High-dimensional datasets show up in numerous fields of study, such as economy, 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. The physical limitations of the display devices and our visual system prevent the direct display and instantaneous recognition of structures with higher dimensions 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 [Ins09, HW13], quality measures [BTK11], clutter reduction [ED07], visual data mining [HG02, Kei02, DOL03], and interactive techniques [BCS96]. High-dimensional aspects of scientific data have also been investigated within the surveys [BH07, KH13]. The surveys [WB94, Cha06, Mun14] focus on the various aspects of visual encoding techniques for multivariate data. 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.

We provide a comprehensive survey of advances in high-dimensional data visualization over the past 15 years, with the following objectives: providing actionable guidance for data practitioners to navigate through a modular view of the recent advances, allowing the creation of new visualizations along the enriched information visualization pipeline, and identifying opportunities for future visualization research.

Our contributions are as follows. First, we propose a categorization of recent advances based on the information visualization (InfoVis) pipeline [CMS99] enriched with customized action-driven classifications (Figure 2, Section 2). We further assess the amount of interplay between user interactions and pipeline-based categorization and put user interactions into a measurable context (Table 1, Section 6). Second, we highlight key contributions of each advancement (Sections 3, Section 4, Section 5). In particular, we provide an extensive survey of visualization techniques derived from topological data analysis (Section 3.5, Section 4.4), a new area of study that provides a multi-scale and coordinatesfree summary of high-dimensional data [Car09]. Furthermore, we connect advances in high-dimensional data visualization with volume rendering and machine learning (Section 7). Finally, we reflect on our categorization with respect to actionable tasks, and identify emerging future directions in subspace analysis, model manipulation, uncertainty quantification, and topological data analysis (Section 8).

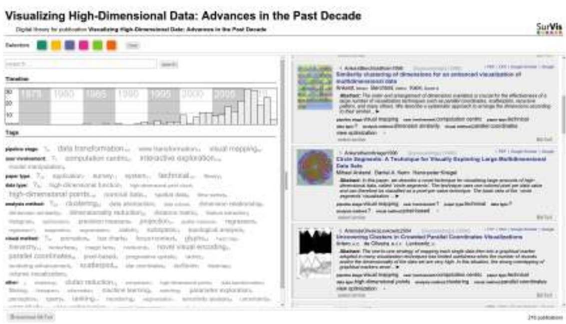

Figure 1: Interactive survey website for paper navigation.

2 Survey Method and Categorization

We conduct a thorough literature review based on relevant works from major visualization venues, namely Visweek, EuroVis, PacificVis, and the journal IEEE Transactions on Visualization and Computer Graphics (TVCG) from the period between 2000 and 2014. To ensure the survey covers the state-of-the-art, we further selectively search through references within the initial set of papers. Beyond the visualization field, we also dedicate 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 (`www.sci.utah.edu/~shusenl/highDimSurvey/website`, based on the SurVis [Bec14] framework; a screen shot is shown in Figure 1) that allows readers to interactively select and filter papers through various tags. However, due to the space limitation, only a subset of the complete list of references (available through the survey website) is mentioned in the paper.

As illustrated in Figure 2, we base our main categorization on the three transformation steps of the information visualization pipeline [CMS99] (and its minor variation in [BTK11]), namely, data transformation, visual mapping, and view transformation. Each category is enriched with novel, customized subcategories. *Data transformation* (Section 3) corresponds to the analysis-centric methods such as dimension reduction, regression, subspace clustering, feature extraction, topological analysis, data sampling, and abstraction. *Visual mapping* (Section 4), the key for most visual encoding tasks, focuses on organizing 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) corresponds to methods focusing on screen space and rendering, including illustrative rendering for various visual structures, as well as screen space measures for reducing clutter or artifacts and highlighting important features.

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

User interactivity is an integral part within each processing step of the pipeline, as illustrated in Figure 2. Based on the amount of user interaction, we can classify all high-dimensional data visualization methods into three categories: computation-centric, interactive exploration, and model manipulation. The distinction between interactive exploration and model manipulation is made to emphasize a particular manipulation paradigm, where the underlying data model is modified based on interaction to reflect user intention. A summary of the interplay between processing steps and interactions is illustrated in Table 1, where user interactions are put into a measurable context. The corresponding details are discussed in Section 6.

3 Data Transformation

We start by describing different types of high-dimensional datasets. We then give an in-depth discussion on the action-driven subcategories centered around typical analysis techniques during data transformation, namely, dimension reduction, clustering (in particular, subspace clustering) and regression analysis. We focus especially on their usages in visualization methods. In addition, we pay special attention to topological data analysis, which is a promising emerging field.

3.1 High-Dimensional Data

We provide an overview of the different aspects of high-dimensional datasets, to define the scope of our discussion and highlight distinct properties of these datasets. Our discussions on different data types are inspired by the book by Munzner [Mun14].

Data Types. In our survey, we limit our exposition to table-based data, and exclude (potentially high-dimensional) graph/network data from the discussion. A high-dimensional dataset is commonly modeled as a point cloud embedded in a high-dimensional space, with the values of attributes corresponding to the coordinates of the points. Based on the underlying model of the data and the analysis and visualization

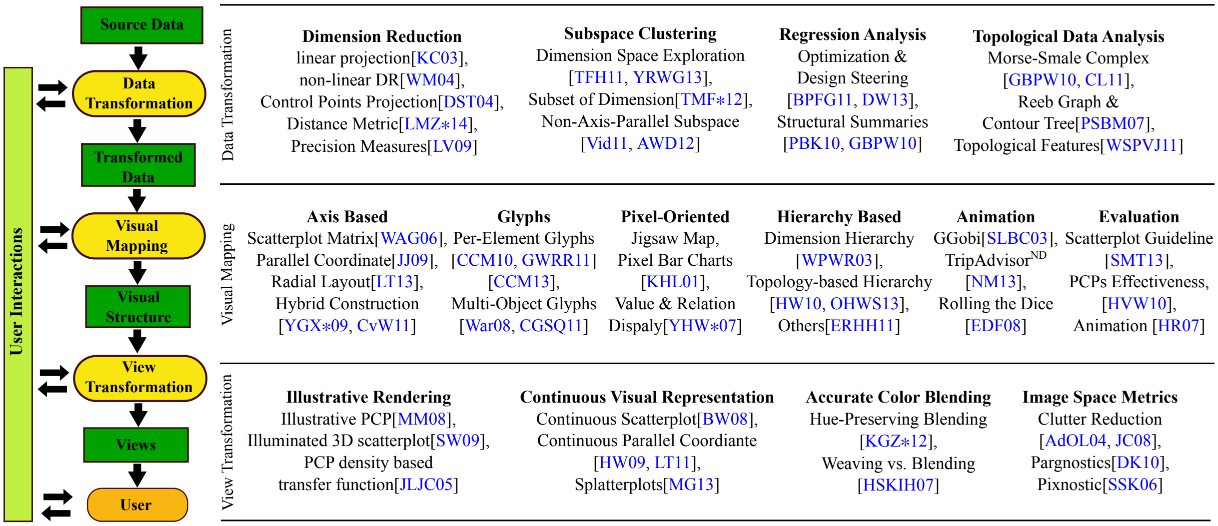

Figure 2: Categorization based on transformation steps within the information visualization pipeline, with customized action-driven subcategories.

goals, the attributes consist of input parameters and output observations, and the data could be modeled as a scalar or vector-valued function (where the function values are based on the output observations) on the point cloud defined by the input parameters. Topological data analysis (Section 3.5) applies to both point cloud data and functions on point cloud data (e.g., [GBPW10, SMC07]), while regression analysis (Section 3.4) typically applies to the latter (e.g., [PBK10]).

Attribute Types. The attribute type (e.g., nominal vs. numerical) can greatly impact the visualization method. In many fields and applications, the value of the attributes is nominal in nature. However, most commonly available high-dimensional data visualization techniques such as scatterplots or parallel coordinate plots are designed to handle numerical values only. When utilizing these methods for visualizing nominal data, information overlapping and visual elements stacking usually exist. One way to address the challenge is mapping the nominal values to numerical values [RRB*04] (e.g. as implemented in the XmdvTool [War94]). Through such a mapping, each axis is used more efficiently and the spacing becomes more meaningful. In the Parallel Sets work [BKH05], 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 [Wea09] allows users to rapidly query nominal values with certain conjunctive relationships through simple interactions. The GPLOM (Generalized Plot Matrix) [IML13] extends the Scatterplot Matrix (SPLOM) to handle nominal data.

Spatiotemporal Data. Some recent advances focus on developing visual encoding that capture the spatiotemporal aspects of high-dimensional data. Visual analysis of the financial time series data is explored in the work by Ziegler et al. [ZJGK10]. The work presented by Tam et al. [TFA*11] studies facial dynamics utilizing the analysis of time-series data in parameter space. Datasets with spatial information such as multivariate volumes [BDSW13] or multi-spectral images [LAK*11] are very common in scientific visualization, and numerous methods have been introduced within the scientific visualization domain, see [BH07, KH13] for comprehensive surveys on these topics. We discuss the intrinsic interconnections between these two areas in Section 7.

3.2 Dimension Reduction

Dimension reduction techniques are key components for many visualization tasks. Existing work either extends the state-of-the-art techniques, or improves upon their capabilities with additional visual aid.

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

PCA [Jol05] is designed to find an orthogonal linear transformation that maximizes the variance of the resulting embedding. PCA can be calculated by an eigendecomposition of the data's covariance matrix or a singular value decomposition of the data matrix. The interactive PCA (iPCA) [JZF*09] 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 recent work by Koren et. al. [KC03] generalizes PCA and LDA by providing a family of flexible linear projections to cope with different kinds of data.

Non-linear Dimension Reduction. There are two distinct groups of techniques in non-linear dimension reduction, under either the metric or non-metric setting. The graph-based techniques are designed to handle *metric* inputs, such as Isomap [TDSL00], Local Linear Embedding (LLE) [RS00], and Laplacian Eigenmap (LE) [BN03], where a neighborhood graph is used to capture local distance proximities and to build a data-driven model of the space.

The other group of techniques address non-metric problems commonly referred to as non-metric MDS or stress-based MDS by capturing non-metric dissimilarities. The fundamental idea behind the non-metric MDS is to minimize the mapping error directly through iterative optimizations. The well-known Shepard-Kruskal algorithm [Kru64] begins by finding a monotonic transformation that maps the non-metric 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. [WM04], where the user is presented with a coarse approximation from partial data. The refinement is on-demand based on user inputs.

Control Points Based Projection. For handling large and complex datasets, the traditional linear or non-linear dimension reductions are limited by their computational efficiency. Some recent developments, e.g., [DST04, PNML08, PEP*11a, JPC*11, PSN10], utilize a two-phases approach, where the control points (anchor points) are projected first, followed by the projection of the rest of the points based on the control points location and local features preservation. 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 desirable 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. [BLBC12] introduce the distance function learning concept, where a new distance metric is calculated from the manipulation of point layouts by an expert user. In [Gle13], 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 [LMZ*14] 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 [LV09] that assesses the ranking discrepancy before and after applying dimension reduction. This technique is then generalized [MLGH13] and used for visualizing dimension reduction quality. A projection precision measure is introduced in [SvLB10], where a local precision score is calculated for each point with a certain neighborhood size. In the distortion-guided exploration work [LWBP14], several distortion measures are proposed for different dimension reduction techniques, where 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. [SSK10] further develop this idea by incorporating the analysis and visualization for better understanding of the localized stress phenomena.

3.3 Subspace Clustering

Clustering is one of the most widely used data-driven analysis methods. Instead of providing an in-depth discussion on all clustering techniques, in this survey, we focus on subspace clustering techniques which have a great impact for understanding and visualizing high-dimensional datasets. Dimension reduction aims to compute one single embedding that best describes the structure of the data. However, this could become ineffective due to the increasing complexity of the data. Alternatively, one could perform subspace clustering, where multiple embeddings can be generated through clustering either the dimensions or the data points, for capturing various aspects of the data.

Dimension Space Exploration. Guided by the user, dimension space exploration methods interactively group relevant dimensions into subsets. Such exploration allows us to better understand their relationships and to identify shared patterns among the dimensions. Turkay et al. introduce a *dual* visual analysis model [TFH11] where both the dimension embedding and point embedding can be explored simultaneously. Their later improvement [TLLH12] 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 [YRWG13] extends a similar concept to allow a recursive exploration of both the dimension space and data space. Several visual encoding methods also rely on the concept of dimension space exploration. These methods are discussed in Section 4.3.

Clustering Subsets of Dimensions. Comparing 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 [CFZ99], originate from the data mining and knowledge discovery community. They introduce some very interesting exploration strategies for high-dimensional datasets, and can be particularly effective when the dimensions are not tightly coupled. The $TripAdvisor^{ND}$ [NM13] 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 [TMF*12] utilizes the SURFING (subspaces relevant for clustering) [BPR*04] algorithm to search the high-dimensional space and automatically identifies a large candidate set of interesting subspaces. For the work presented by Ferdosi et al. [FBT*10], 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 clustering the dimensions, which essentially creates axis-aligned linear subspaces, identifying non-axis-aligned subspaces is a more flexible alternative. Projection Pursuit [FT74] is one of the earliest works aimed at automatically identifying the interesting non-axis-aligned subspaces. Projections are considered to be more interesting when they deviate more from a normal distribution. Some advances have been made in the machine learning community to perform non-axis-aligned subspace clustering [Vid11]. Instead of clustering the dimensions, the points are grouped together for sharing similar linear subspaces. In particular, we assume 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.

For very high-dimensional data, the subspace finding algorithms typically have a relatively high computational complexity. By utilizing random projection, Anand et al. [AWD12] introduce an efficient subspace finding algorithm for data with thousands of dimensions. It 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 in high dimension is an extensive and active field of research in its own right. We make no attempt to survey the entire area, but rather focus 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 where, in addition to a multivariate input space, the user has one or several output or response variables they want to explore (e.g., [BPFG11,TWSM*11]), where the results require a qualitative examination, or are used to inform decisions.

HyperMoVal [PBK10] is a software system used for validating regression models against actual data. It uses support vector regression (SVR) [SS04b] 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 their regression to an acceptable level of accuracy. Berger et al. [BPFG11] utilize two different 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 are encoded around the focal point.

Tuner [TWSM*11] begins as an automated adaptive sampling algorithm where a sparse sampling of the parameter space is refined by building a Gaussian Process Model (GPM) (see [RW06] for a good overview) 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. [DW13] 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. [RPH08] and Gerber et al. [GBPW10]. Both methods summarize the structures of the data via skeleton representations. Reddy et al. [RPH08] 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 principle curves [HS89] which go through the *medial-axis* of the data. HDViz [GBPW10], 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 post-processing 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 (TDA), a new field of study (see [Zom05, BDFF*08, EH08, EH10, Car09, Ghr08] for seminal works and surveys), has provided efficient and reliable feature-driven analysis and visualization capabilities. Specifically, the construction of topological structures [Ree46, Sma61] from scalar functions on point clouds (e.g., Morse-Smale complexes, contour trees, and Reeb graphs) as "summaries" over data is at the core of such TDA methods. 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 (Figure 3). They both provide meaningful abstractions of the high-dimensional data, which reduces the amount of data needed to be processed or stored; and they utilize sophisticated hierarchical representations capturing features at multiple scales, which enables progressive simplifications of features differentiating small and large scale structures in the data.

Morse-Smale Complexes. The Morse-Smale complex (MSC) [EHNP03, EHZ03] 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 of 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, merging nearby clusters by redirecting the gradient flow. MSCs have been shown to be effective in identifying, ordering, and selectively removing features of large-scale data in scientific visualizations (e.g., [BEHP04, GBPH08, GNP*05]).

HDViz [GBPW10] 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. Correa et al. [CL11] suggest that by considering a different type of neighborhood structure, we can improve the accuracy in the extracted topology compared to those obtained within HDViz.

Reeb Graphs and Contour Trees. The Reeb graph of a real-valued function describes the connectivity of its level sets. A contour tree is a special case of Reeb graph that arises in simply-connected domains. 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., [NLC11, SMC07]).

Efficient algorithms for computing the contour tree [CSA03] and Reeb graph [PSBM07] in arbitrary dimensions have been developed. A generalization of the contour tree has been introduced by Carr et al. [CD14, DCK*12] called the joint contour net (JCN), which allows for the analysis of multi-field data.

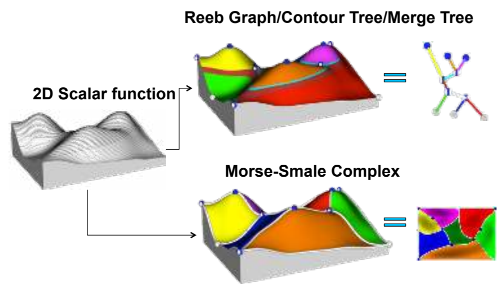

Figure 3: Contour- and gradient-based topological structure of a 2D scalar function.

Other Topological Features. Ghrist [Ghr08] and Carlsson [Car09] both offer several applications of TDA and in particular highlight the topological theory used in a study of statistics of natural images [LPM03]. Mapper [SMC07] 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. Wang et al. [WSPVJ11] utilize TDA techniques developed by Silva et al. [dSMVJ09] to recover important structures in high-dimensional data containing non-trivial topology. Specifically, they are interested in high-dimensional branching and circular structures. The circle-valued coordinate functions are constructed to represent such features. Subsequently, they perform dimension reduction on the data while ensuring such structures are visually preserved.

4 Visual Mapping

Visual mapping plays an essential role in converting the analysis result or the original dataset into visual structures based on various visual encodings. Here, we divide the approaches based on their structural patterns, compositions, and movements (i.e., animations). In addition, the methods that evaluate the effectiveness of visual encoding are also discussed.

4.1 Axis-Based Methods

Axis-based methods refer to visual mappings where element relationships are expressed through axes representing the dimensions/variables. These methods include some of the most well-known 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 the 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 semi-automatically identifying more interesting plots.

Scagnostics are a set of measures designed for identifying interesting plots originally introduced by John W. Tukey. The recent works of Wilkinson et al. [WAG05, WAG06] extend the concept to include nine measures capturing properties such as outliers, shape, trend, and density. In addition, they improve the computational efficiency by using graph-theoretic measures. Scagnostics is also extended to handle time series data [DAW13]. Guo [Guo03] 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 [SS04a, SS06] 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 [SNLH09] as a quality metric for 2D scatterplots. The class consistency measure is defined by the distance to the class's center or entropies of the spatial distributions of classes. Tatu et al. [TAE*09] 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. [LAE*12] introduces 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, where only bivariate relationships can be directly expressed, the Parallel Coordinate Plot (PCP) [Ins09, ID91] allows patterns that highlight multivariate relations to be revealed by showing all the axes at once (typically, in a vertical layout). However, due to the linear ordering of the PCP axes, for a given n-dimensional dataset, there are $n!$ permutations of the ordering of the axes. Each of the orderings highlights certain aspects of the high-dimensional structure. Therefore, one of the significant challenges when dealing with PCPs is determining an appropriate order of the axes. In addition, as the number of points increases, the line density in the PCP increases dramatically, which can lead to overplotting and visual clutter thus hindering the discovery of patterns.

A few methods have proposed metrics for ordering the axes automatically. Tatu et al. [TAE*09] introduce PCP ranking methods 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 [FR11] that is applicable for both PCPs and SPLOMs utilizing the subspace analysis method from their earlier work [FBT*10] discussed in the Section 3.3. Johansson and Johansson [JJ09] 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. Hurley et al. utilize Eulerian tours and Hamiltonian decompositions of complete graphs, which represent the relationship between the dimensions, in their recent work [HO10] to address the axis ordering challenge.

Clutter reduction is another important aspect in PCPs, especially for large point counts. Peng et al. [PWR04] were able to reduce clutter for both SPLOMs and PCPs without altering the information content simply by reordering the dimensions. A focus+context visualization scheme can also be used for reducing the clutter and highlighting the essential features in the PCP [NH06]. In this context, the overview captures both the outliers and the trends in the dataset. The outliers are indicated by single lines, and the trends that capture the overall relationship between axes are approximated by polygon strips. The selected data items are emphasized through visual highlighting. In addition, several of the clutter reduction methods employing screen space measures are discussed in detail in Section 5.4.

Finally, many visual encoding improvements exist for PCPs. Progressive PCPs [RZH12] demonstrate the power of a progressive refinement scheme for enhancing the ability of PCPs to handle large datasets. In the work of Dang et al. [DWA10], 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.

Radial Layout. The star coordinate plot [Kan00], also referred to as a bi-plot [HGM*97], 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 where substantial distortion occurs. Lehmann et al. extend the star coordinate concept with an orthographic constraint [LT13] to restrict the generated projection to be orthographic, which better preserves the structure of the original dataset.

Similar to the star coordinates, Radviz [HGM*97] 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. [AEL*10] devise a RadViz quality measure allowing automatic optimization of the dimensional anchor layout.

DataMeadow [EST07] introduces a radial visual encoding named DataRoses, which is represented as a PCP laid out radially as opposed to linearly. Lastly, PolarEyez [JN02] introduces a focus+context visualization where 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.

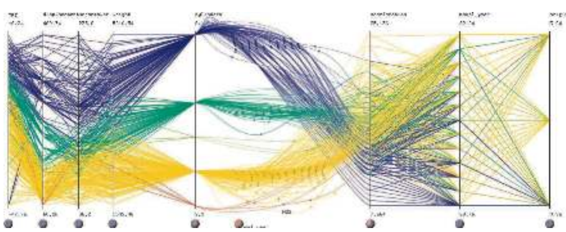

Figure 4: Scattering points in parallel coordinates by Yuan et al. [YGX*09].

Hybrid Construction. The axis-based methods can also be combined to create new visualizations. The scattering points in parallel coordinate work [YGX*09] (Figure 4) embeds a MDS plot between a pair of PCP axes. The flexible linked axes work [CvW11] 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 [FCI05] provides a way to "unfold" the overlapped values in the PCP axis in 3D space. In this work [FCI05], each axis in the PCP is replaced with a star glyph that represents the values 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, there is a number of visual representations that derive from the the well-known visual mappings. Angular histograms [GPL*11] introduced a novel visual encoding that improves the scalability of PCPs by overcome the overplotting issue. The tiled PCP [CMR07] 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

Chernoff faces [Che73] 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 can be visually encoded into scatterplots [CCM09, CCM10, GWRR11, CCM13].

Correa et al. [CCM09] aimed at incorporating uncertainty information into PCA projections and k-means clustering and accomplished this goal by augmenting scatterplots with tornado plots. Together these glyphs encode uncertainty and partial derivative information. The idea of mapping sensitivity information to a line segment through each data point has been extended in their later work [CCM10] with the introduction of the flow-based scatterplot (FBS) that highlights functional relationships between inputs and outputs. The works by Guo et al. [GWRR11] and Chan et al. [CCM13] attempt to provide more than a single partial derivative information into their scatterplots by experimenting with different glyph shapes such as star plots among others. [GWRR11] also uses a bar chart similar to the tornado plot used in [CCM09], and [CCM13] provides two other interpretations. The first is a generalization of the FBS called the generalized sensitivity scatterplot (GSS). By using orthogonal regression, GSSs can represent the partial derivative of any variable with respect to any other variable. The other is a fan glyph that works similarly to the star glyph, allowing for viewing multiple partial derivatives, but rather than displaying magnitude as in the star glyph, the fan glyph highlights the direction of each partial derivative, since all line segments are normalized in length.

The methods described above all deal with encoding extra information per data point into glyphs, but the DICON system [CGSQ11] attempts to show the trend of data within a collection of data points by visually encoding statistical information about the set of points being represented. DICON uses dynamic icons based on treemap visualization to encode clusters of data into separate icons, and allows the user to interactively merge, split, filter, regroup, and highlight clusters or data within clusters. Due to the interactive nature, the authors have developed a stabilized Voronoi layout that allows data within the treemap to maintain spatial coherence as the user edits the clusters. They further encode skew and kurtosis into the shape of the icon before applying the Voronoi algorithm, thus allowing for statistical details to be presented.

Finally, Ward [War08] gives a thorough, practical treatment of 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 [KK94, AKpK96]. VisDB [KK94] 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 [AKpK96] 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 [KHL*01]. Pixel bar charts first separate data into separate bars based on one dimension or attribute, and it 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 [Wat05], 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 [YPS*04, YHW*07] combine the recursive pattern displays [KKA95] with MDS in order to lay out the separate subwindows such that similar dimensions are placed closer together. A latter iteration [YHW*07] enhances the work by providing more robust visualizations including jigsaw maps, scatterplot glyphs, and a novel concept known as the Rainfall metaphor geared at establishing the relationship of all dimensions to a single dimension of interest.

4.4 Hierarchy-Based Approaches

For visualizing high-dimensional datasets, hierarchical visual representations are used to capture dimensional relationships, represent contour tree structure, and provide new visual encodings for representing high-dimensional structures.

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 [WPWR03] based on dimension similarity. The dimension hierarchy is represented and navigated by a multiple ring structure (InterRing [YWR02]), where the innermost-ring represents the coarsest level in the hierarchy.

Topology-Based Hierarchies. In Section 3.5, 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.

Various visual metaphors have been designed for contour trees [PCMS09, WBP12]. In particular, variations of topological landscapes have been proposed [BMW*12, DBW*12, HW10, OHJS10, OHJ*11, WBP07]. These visual metaphors have, or potentially have, capabilities for the visualization of high-dimensional datasets. In particular, Weber et al. [WBP07] have presented such a metaphor for visually mapping the contour tree of high-dimensional functions to a 2D terrain where the relative size, volume, and nesting of the topological features are preserved. Harvey and Wang [HW10] have extended this work by computing all possible planar landscapes and 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. [OHJS10, OHJ*11] 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. [OHWS13] continued this line of work by creating a linked view software system including user interactions into 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 either persistence, cluster size, or cluster stability, thus adjusting the placement of features in the topological landscape.

Other Hierarchical Structures. In the structure-based brushes work [FWR00], a data hierarchy is constructed to be visualized by both a PCP and a treemap [Shn92], allowing users to navigate among different levels-of-detail and select the feature(s) of interest. The structure decomposition tree [ERHH11] 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 [KS02] for visualizing complex hierarchical graphs in a focus+context manner for visual data mining tasks.

4.5 Animation

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 [SLBC03] provides a mechanism for calculating a continuous linear projection transition between any pair of linear projections based on the principal angles between them. In the Rolling the Dice work [EDF08], 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. Rnav-Graph [WO11] constructs a graph connecting a number of interesting scatterplots. A smooth animation is generated between all scatterplots that are connected by an edge. The $TripAdvisor^{ND}$ [NM13] 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 [SMT13], including 2D scatterplot, interactive 3D scatterplot, 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. The efficacy of several PCP variants for cluster identification has been studied in [HVW10]. The 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 [YGX*09], a number of seemingly valid improvements do not result in significant performance gains for cluster identification tasks. Heer et al. investigate the animated transition effectiveness between statistic graphs [HR07] such as bar charts, pie charts, and scatterplots. Their results reveal that animated transitions, when used appropriately, can significantly improve graphical perception.

5 View Transformation

View transformations dictate what we ultimately see on the screen. As pointed out by Bertini et al. [BTK11], 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 that focus on achieving a specific visual style by applying custom-designed rendering algorithms. The illustrative PCPs work [MM08] provides a set of artistic style rendering techniques for enhancing parallel coordinate visualization. Some of the rendering techniques include spline-based edge bundling, opacity-based hints to convey cluster density, and shading effects to illustrate local line density. Illuminated scatterplots [SW09] (Figure 5) 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. [JLJC05] 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 features.

Illustrative rendering techniques are also used for highlighting the focused areas, such as the well-known Table-Lens approach [RC94] for visualizing large tables. Such a magic lens based approach permits fast exploration of an area of interest without presenting all the details, therefore, reduces clutter in the view. MoleView [HTE11], 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 data to maintain context. A survey on the distortion-oriented magic lens techniques is presented by Leung and Apperley [LA94].

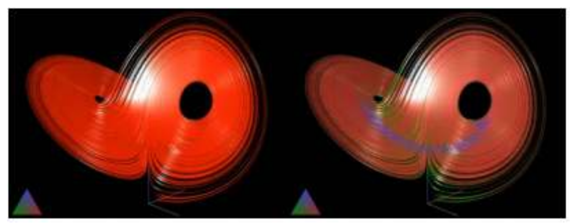

Figure 5: Illuminated 3D scatterplot by Sanftmann et al. [SW09].

5.2 Continuous Visual Representation

For most high-dimensional visualization techniques, a discrete visual representation is assumed since each element corresponds to a 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 [BW08] presents a mathematical model for constructing a continuous scatterplot. The follow-up work [BW09] introduces an adaptive rendering extension for continuous scatterplots increasing the rendering efficiency. This concept is extended to create continuous PCPs [HW09] based on the point and line duality between scatterplots and parallel coordinates. The authors propose a mathematical model that maps density from a continuous scatterplot [BW08] to parallel coordinates. Lehmann et al. introduce a feature detection algorithm design for continuous PCPs [LT11].

Clutter caused by overlapping in PCPs and scatterplots occludes data distribution and outliers. In the splatterplot work [MG13], 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 data points into an abstract contour representation 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 by [ZCQ*09], 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 semi-transparent objects, color blending methods have a significant impact on the perception of order and structure.

As stated in the Hue-Preserving Color-Blending work [KGZ*12], 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 a hue-preserving blending. This model can be applied to high-dimensional visualization techniques such as illustrative PCPs [MM08], where a depth ordering clue is better preserved. In the Weaving vs. Blending work [HSKIH07], 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 [AdOL04] 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 [JC08] to address the challenge of very large datasets. The metric is based on distance transformation, and the computation is carried out on the GPU for interactive performance.

Pargnostics [DK10], a portmanteau for parallel coordinates and diagnostics (similar to Scagnostics [WAG05]), is a set of screen space measures for identifying distance patterns among pairs of axes in PCPs. The metrics include line crossings, crossing angles, convergence, and over-plotting. For each of the metrics, the system provides ranked views for pairs of axes, allowing the user to guide exploration and visualization. Pixnostic [SSK06] is an image space based quality metric for ranking interestingness for pixel based (Section 4.3) visualization such as Pixel Bar Chars [KHL*01].

6 User Interaction

As illustrated in Figure 2, interaction is integrated with each of the processing steps. An alternative subcategorization for each of the processing steps based on the amount of user interaction is shown in Table 1. In this categorization, each step is further divided into computation-centric approaches, interactive exploration, and model manipulation. In both of the recent surveys [MPG*14, TJHH14] on the user interaction in visualization applications, the level of integration between the computation and visualization (indicate user interaction) is used for classifying the methods. In many ways, their classifications are aligned with our proposed approach, with the distinction that our discussion is directly connected to the information visualization pipeline.

6.1 Computation-Centric Approaches

Computation-centric 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 [RS00, MRC02, KC03, WM04], subspace analysis [CFZ99, TMF*12, FBT*10, AWD12], regression analysis [BPFG11, BPFG11], quality metric based ranking [WAG05, TAE*09], etc. Computation-centric approaches exist at each of the processing steps, but are most concentrated in the data transformation step.

6.2 Interactive Exploration

Interactive methods navigate, query, and filter the existing model interactively for more effective visual communication. In this section, we focus only on representative methods where the interactive exploration mechanism is their key contribution.

In the data transformation step, the interactive exploration scheme allows users to guide progressive dimension reduction, where a partial result is presented upon request [WM04]. In works by Turkay et al. [TFH11, TLLH12] and Yuan et al. [YRWG13], a subset of dimensions is interactively selected and explored in dimension space.

In the visual mapping step, there are large number of methods focused on interactive exploration and querying the high-dimensional dataset. Such methods play an important role in the knowledge Discovery in Databases (KDD) process, where the term *visual data mining* [KK96, Kei02, DOL03] is used to describe these applications. Interactive filtering, zooming, distortion, linking and brushing, or a combination of them have been adopted to include the user as part of the exploring and querying process. Polaris [STH02] is a visual query and analysis system designed for relational databases. This system is later developed into the well-known commercial product Tableau. Stolte et al. introduce an approach for zooming along one or more dimensions for multi-scale exploration by traversing a graph [STH03]. In this system, relational queries can be defined by visual specifications allowing fast incremental development and intuitive understanding of the data. Hao et al. have introduced the Intelligent Visual Analytics Queries [HDK*07]. Their approach utilizes correlation and similarity measurements for mining data relationships. We believe new research directions could stem from visual

	Data Transformation	Visual Mapping	View Transformation
Computation-Centric	dimension reduction [WM04], subspace finding [TMF*12], regression analysis	automatic parallel coordinate axis reordering [PWR04], scatterplot ranking [WAG05]	quality metrics in image space [JC08], continuous visual representation [BW08]
Interactive Exploration	interactive, progressive dimension reduction [WM04], dimension space exploration [TFH11]	visual querying and filtering [SVW10], animated transition [SLBC03]	interactive magic lens effects [HTE11], illuminated 3D scatterplot [SW09]
Model Manipulation	user-guided embedding manipulation [LWBP14], control point based projection [JPC*11]	distance function learning [BLBC12, Gle13], visual to parameter interaction [HBM*13]	PCP transfer function [JLJC05], inverse projection extrapolation [PdSABD*12]

Table 1: The transformation pipelines intertwine with user interaction. The subcategorizing is based on the different levels of user involvement.

data mining and visual queries. The Select and Slice Table [SVW10] allows users to study the relationships between data subsets and the semantic zone (user-defined areas of interest). The semantic zones are arranged along one axis of the table, while the data subsets are arranged along the other axis. In addition, the method enables the combination and manipulation of the semantic zones for further exploration. More recent works [GLG*13, GGL*14] by Gratzl et al. introduce some very interesting interactive methods for rankings multi-attributes and explore subsets of tabular datasets.

Both of the works of Poco et al. [PEP*11b] and Sanftmann and Weiskopf [SW12] present methods for navigating a 3D projection. However, their approaches are quite different. The method introduced by Poco et al. [PEP*11b] focuses on enhancing the visual encoding and exploration usability of a 3D projection calculated by the Least-Square Projection [PNML08] algorithm. On the other hand, Sanftmann and Weiskopf [SW12] present an interpolation scheme for generating 3D rigid body rotations between a pair of 3D axis-aligned scatterplots that share a common axis.

In the view transformation step, interactivity is inherent in both the magic lens based methods [HTE11,LA94], and illuminated 3D scatterplots [SW09] (discussed in Section 5.1).

6.3 Model Manipulation

Model manipulation 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 [BLBC12], 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. Hu et al. present a method [HBM*13] for improving the translation of user interaction to algorithm input (visual to parameter interaction) for distance learning scenarios. The explainers [Gle13] are projection functions created from a set of user-defined annotations.

The control point based projection methods [DST04, PNML08, PEP*11a, JPC*11, PSN10] update the overall projection result based on user manipulation of the control points. In the iLAMP method [PdSABD*12], inverse projection extrapolation is used for generating synthetic multidimensional data out of existing projections for parameter space exploration. In the Local Clustering Operation work [GXWY10], the visual structure is modified in PCPs through user-guided deformation operators. Finally, Liu et al. [LWBP14] allow for direct manipulation of the dimension reduction embedding to resolve structural ambiguities. The interactively updated distortion measure is used for feedback during manipulation.

7 Connections with Related Fields

We investigate the connections between recent advances in high-dimensional data visualization and related fields in the hope of inspiring new research directions.

7.1 Multivariate Volume Visualization

Multivariate volume visualization and high-dimensional visualization are often studied under different contexts: the former is normally considered as scientific visualization research [BH07,KH13], while the latter is mostly studied from the perspective of information visualization and visual analytics. In addition, they focus on different kinds of data and attempt to accomplish distinct goals.

Despite the differences, recent advances in both areas have shown that they share a number of fundamental techniques and principles. Standard high-dimensional data visualization techniques, such as PCPs, scatterplots, and dimension reduction, have found their way into the multivariate volume visualization literature. For example, the scattering points in parallel coordinates work [YGX*09] is adopted by [GXY12] as a design space for multivariate volume transfer functions. In the work of Liu et al. [LWT*14a], dynamic projection and subspace analysis are utilized for exploring the high-dimensional parameter space of volumetric data. We believe useful and interesting techniques may be developed by sharing ideas and discovering new connections between these two fields.

7.2 Machine Learning

Machine learning algorithms under many situations have been treated as "black box" approaches, and the parameter tuning process can be tedious and unpredictable. To resolve such a challenge, several visualization approaches have been introduced to aid the understanding of the various machine learning algorithms. Tzeng et al. present a visualization system that helps users design neural networks more efficiently [TM05]. The works of Teoh and Ma [TM03] and van den Elzen and van Wijk [vdEvW11] investigate visualization methods for interactively constructing and analyzing decision trees. Visualization has also been used to aid model validation [Rd00, MW10]. Numerous challenges for understanding machine learning algorithms coincide with high-dimensional visualization. We believe high-dimensional visualization will play an important role in designing, tuning, and validating machine learning algorithms.

8 Reflections and Future Directions

One of our primary objectives in presenting this survey is to provide actionable guidance for data practitioners to navigate through a modular view of the recent advances. To do so, we provide a categorization of recent works along an enriched information visualization pipeline. We reflect on the chosen categories and subcategories (as described briefly in Section 2) and describe on a high level how they provide actionable guidance. To allow the creation of new visualizations along the pipeline, one should think beyond data tasks to be performed in any single stage, and focus on understanding how results from one stage could be utilized most effectively in the remaining stages. We argue that the subcategories discussed during each pipeline stage correspond to sets of actionable items or toolsets that the data practitioner could choose from and rely upon. The combinations of techniques they chose to apply are largely data-driven and application-dependent. Nevertheless, the techniques surveyed following our categorization aim to provide a modular view during the design process.

We now discuss the challenges addressed by the techniques surveyed in the paper, and those that remain to be tackled. Our discussion is partially inspired by Donoho's AMS lecture [Don00] where he discusses the curses and blessings of dimensionality when it comes to high-dimensional data analysis.

Data analysis, falls under the *data transformation* stage within our categorization. Some of the surveyed, standard data analysis tasks are widely applicable for studying various aspects of high-dimensional data: dimension reduction for feature selection and extraction; clustering for exploratory data mining and classification; regression for relationship inference and prediction. However, we identify several different directions in which we expect to see further progress, namely: robust analysis and data de-noising; multiscale analysis; data skeletonization; and high-dimensional approximations. First, more advanced regression techniques could be developed that are robust to noise and outliers, in particular, a new class of regression techniques inspired by geometric and topological intuititions (e.g., [GBPW10]). Second, topological data analysis has built-in capabilities in separating features from noise at multi-scales; such a multiscale notion is expected to be transferrable to a larger class of analysis techniques. Third, developing frameworks to extract as well as to simplify "skeletons" from high-dimensional data can be extremely useful for visual data abstraction and exploration (e.g., [SMC07]). Finally, as pointed out by Donoho [Don00], perhaps there exists new notions of high-dimensional approximation theory, where we make different regularity assumptions and obtain a very different picture in approximating high-dimensional functions. Approximating the Morse-Smale complex in high dimension is considered such an example.

During *visual mapping*, our surveyed techniques convert the analysis result into visual structures with various visual encodings. Development of new analysis results, for example, new approximations of high-dimensional structures, would inevitably lead to new visual metaphors (e.g., in the case of topological landscape [HW10, DBW*12]). Under *visual mapping* and *view transformation*, we also see various methods aimed at summarizing trends in data, such as glyph representations, edge bundling in PCPs, splatting as presented in splatterplots [MG13] and PCP-based splatters [ZCQ*09], and hierarchical approaches. These could be further enhanced with new data skeletonization techniques.

Finally, we identify a few opportunities for future visualization research and discuss them in detail.

Subspace Clustering. Finding interesting projections (views) has been an active and important research area for visualizing high-dimensional data. The motivation behind the various view selection schemes can be traced back to much earlier work such as projection pursuit [FT74].

Along a similar line of research, scatterplot ranking methods [SS04a, WAG06, TAE*09] are introduced to automatically identify the interesting scatterplots. However, a scatterplot matrix captures only limited bivariate relationships. Subspace selection methods [CFZ99, BPR*04], originally developed in the data mining community, have recently been adapted for high-dimensional data visualization [FBT*10, TMF*12] to capture more complicated multivariate structures. Despite the added flexibility, the search is still limited to axis-aligned subspaces. Recent advances in machine learning, such as subspace clustering (e.g., [Vid11]), assume the high-dimensional dataset can be represented by a mixture of low-dimensional linear subspaces with mixed dimensions. Such methods produce non-axis-aligned subspaces, which work well for datasets where different dimensions are closely related. In addition, instead of capturing a single linear subspace, they can approximate non-linear structures by fitting together multiple linear subspaces.

We believe exploring various (non-axis-aligned) subspace clustering methods will lead to new developments in high-dimensional view selection techniques (e.g., some of the recent work by the authors [LWT*14a, LWT*14b]).

Model Manipulation. We have seen an emerging user interaction paradigm, referred to as model manipulation [BLBC12, PdSABD*12, Gle13, HBM*13] in this survey. What differentiates the model manipulation interaction from other types of interaction is the change of the underlying data model to reflect user intention. These model manipulation based methods allow users to easily transfer their domain knowledge into the exploratory analysis process, allowing for effective analysis and visualization. However, since such interactive manipulations give users an enormous amount of freedom, one of the main challenges in model manipulation is to understand whether or not the manipulation faithfully conveys the user intention. Rigorous validation between the user intended operations and manipulation outcomes is essential for evaluating the effectiveness and usability of these methods.

Uncertainty Quantification. Along with the large-scale and high dimensionality of the data, information pertaining to uncertainty is becoming increasingly available and important. The addition of uncertainty information within visualizations has been deemed a top research problem in scientific visualization [Joh04], due to the greater availability of this information from simulation and quantification, and the importance of understanding data quality, confidence, and error issues when interpreting scientific results. Some recent works in high-dimensional data visualization have focused on analyzing the uncertainty stemming from the input data or with respect to the accuracy of a fitted model (see Section 3.4 and [ZSWR06]). We believe the extensions and generalizations of existing uncertainty visualization capabilities (e.g., [DKLP02, PWB*09, SZD*10]) to high-dimensional data is one of the important future directions.

Another interesting aspect of uncertainty quantification is based on uncertainty-aware visual analytics discussed by Correa et al. [CCM09], and further explored by Liu et al. [LWBP14] and Schreck et al. [SvLB10], where the uncertainty (e.g., bias and distortions) arises from the *Data transformation* step. The work by Correa et al. [CCM09] measures the uncertainty introduced by three common *Data transformation* techniques; and the works of Liu et al. [LWBP14] and Schreck et al. [SvLB10] quantifies the amount of distortion for projection techniques. While these methods apply to the uncertainty stemming from the *Data transformation* step, more work can be done to define measures of uncertainty associated with the two latter processing steps in the visualization pipeline, namely *Visual mapping* and *View transformation*.

Topological Data Analysis and Visualization. Another important and interesting recent advance is the introduction of TDA to visualization (e.g., [GBPW10, WSPVJ11, DCK*12]). TDA provides an interesting alternative for capturing the structure in high-dimensional data. Since topological structures are typically scale-invariant, designing meaningful and effective visual encodings that capture their inherent properties is essential for future development. Approximation algorithms exist for computing topological structures in high dimensions; therefore, it is important to strike a balance between speed and accuracy, and to convey appropriately the approximation error in the visualization. Some initial work has been done to provide bounds or estimations on the accuracy of these approximated models (e.g., [GBPW10, CL11, TFO09]).

Other Directions. Finally, as discussed in Section 7, fields such as multivariate volume visualization and machine learning share a number of common research problems with high-dimensional data visualization. Finding connections and sharing ideas among these related topics will likely not only yield interesting future research directions, but also help resolve many challenges in high-dimensional data visualization.

Acknowledgments

The first two authors contributed equally to this work. 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 NSF 0904631, DE-EE0004449, DE-NA0002375, DE-SC0007446, DE-SC0010498, NSG IIS-1045032, NSF EFT ACI-0906379, DOE/NEUP 120341, DOE/Codesign P01180734.

References

[AdOL04] ARTERO A., DE OLIVEIRA M., LEVKOWITZ H.: Uncovering clusters in crowded parallel coordinates visualizations. In *IEEE Symposium on Information Visualization* (2004), pp. 81–88. 11

[AEL*10] ALBUQUERQUE G., EISEMANN M., LEHMANN D. J., THEISEL H., MAGNOR M.: Improving the visual analysis of high-dimensional datasets using quality measures. In *IEEE Symposium on Visual Analytics Science and Technology* (2010), IEEE, pp. 19–26. 8

[AKpK96] ANKERST M., KEIM D. A., PETER KRIEGEL H.: Circle segments: A technique for visually exploring large multidimensional data sets. In *Proceedings of IEEE Visualization, Hot Topic Session.* (1996). 9

[AWD12] ANAND A., WILKINSON L., DANG T. N.: Visual pattern discovery using random projections. In *IEEE Conference on Visual Analytics Science and Technology* (2012), IEEE, pp. 43–52. 5, 11

[BCS96] BUJA A., COOK D., SWAYNE D. F.: Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics 5*, 1 (1996), pp. 78–99. 1

[BDFF*08] BIASOTTI S., DE FLORIANI L., FALCIDIENO B., FROSINI P., GIORGI D., LANDI C., PAPALEO L., SPAGNUOLO M.: Describing shapes by geometrical-topological properties of real functions. *ACM Computing Surveys 40*, 4 (2008), 12:1–12:87. 6

[BDSW13] BISWAS A., DUTTA S., SHEN H.-W., WOODRING J.: An information-aware framework for exploring multivariate data sets. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2683–2692. 3

[Bec14] BECK F.: Survis. https://github.com/fabian-beck/survis, 2014. 2

[BEHP04] BREMER P.-T., EDELSBRUNNER H., HAMANN B., PASCUCCI V.: A topological hierarchy for functions on triangulated surfaces. *IEEE Transactions on Visualization and Computer Graphics 10*, 385-396 (2004). 6

[BH07] BÜRGER R., HAUSER H.: Visualization of multi-variate scientific data. *EuroGraphics State of the Art Reports (STARs)* (2007), 117–134. 1, 3, 12

[BKH05] BENDIX F., KOSARA R., HAUSER H.: Parallel sets: visual analysis of categorical data. In *IEEE Symposium on Information Visualization* (2005), pp. 133–140. 3

[BLBC12] BROWN E. T., LIU J., BRODLEY C. E., CHANG R.: Dis-function: Learning distance functions interactively. In *IEEE Conference on Visual Analytics Science and Technology* (2012), IEEE, pp. 83–92. 4, 12, 14

[BMW*12] BEKETAYEV K., MOROZOV D., WEBER G. H., ABZHANOV A., HAMANN. B.: Geometry–preserving topological landscapes. In *Proceedings of the Workshop at SIGGRAPH Asia* (2012), pp. 155–160. 9

[BN03] BELKIN M., NIYOGI P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation 15*, 6 (2003), 1373–1396. 4

[BPFG11] BERGER W., PIRINGER H., FILZMOSER P., GRÖLLER E.: Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction. *Computer Graphics Forum 30*, 3 (2011), 911–920. 5, 11

[BPR*04] BAUMGARTNER C., PLANT C., RAILING K., KRIEGEL H.-P., KROGER P.: Subspace selection for clustering high-dimensional data. In *Fourth IEEE International Conference on Data Mining* (2004), IEEE, pp. 11–18. 5, 13

[BTK11] BERTINI E., TATU A., KEIM D.: Quality metrics in high-dimensional data visualization: an overview and systematization. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2203–2212. 1, 2, 10

[BW08] BACHTHALER S., WEISKOPF D.: Continuous scatterplots. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1428–1435. 10, 12

[BW09] BACHTHALER S., WEISKOPF D.: Efficient and adaptive rendering of 2-d continuous scatterplots. *Computer Graphics Forum 28*, 3 (2009), 743–750. 10

[Car09] CARLSSON G.: Topology and data. *Bullentin of the American Mathematical Society 46*, 2 (2009), 255–308. 2, 6

[CCM09] CORREA C., CHAN Y.-H., MA K.-L.: A framework for uncertainty-aware visual analytics. In *IEEE Symposium on Visual Analytics Science and Technology* (2009), pp. 51–58. 8, 14

[CCM10] CHAN Y.-H., CORREA C., MA K.-L.: Flow-based scatterplots for sensitivity analysis. In *IEEE Symposium on Visual Analytics Science and Technology* (2010), IEEE, pp. 43–50. 8

[CCM13] CHAN Y.-H., CORREA C., MA K.-L.: The generalized sensitivity scatterplot. *IEEE Transactions on Visualization and Computer Graphics 19*, 10 (2013), 1768–1781. 8

[CD14] CARR H., DUKE D.: Joint contour nets. *IEEE Transactions on Visualization and Computer Graphics 20*, 8 (2014), 1100–1113. 6

[CFZ99] CHENG C.-H., FU A. W., ZHANG Y.: Entropy-based subspace clustering for mining numerical data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (1999), ACM, pp. 84–93. 5, 11, 13

[CGSQ11] CAO N., GOTZ D., SUN J., QU H.: Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2581–2590. 8

[Cha06] CHAN W. W.-Y.: A survey on multivariate data visualization. *Department of Computer Science and Engineering. Hong Kong University of Science and Technology 8*, 6 (2006), 1–29. 1

[Che73] CHERNOFF H.: The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association 68*, 342 (1973), 361–368. 8

[CL11] CORREA C., LINDSTROM P.: Towards robust topology of sparsely sampled data. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1852–1861. 6, 14

[CMR07] CAAT M., MAURITS N., ROERDINK J.: Design and evaluation of tiled parallel coordinates visualization of multichannel eeg data. *IEEE Transactions on Visualization and Computer Graphics 13*, 1 (2007), 70–79. 8

[CMS99] CARD S. K., MACKINLAY J. D., SHNEIDERMAN B.: *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999. 1, 2

[CSA03] CARR H., SNOEYINK J., AXEN U.: Computing contour trees in all dimensions. *Computational Geometry 24*, 2 (2003), 75 – 94. Special Issue on the Fourth CGC Workshop on Computational Geometry. 6

[CvW11] CLAESSEN J., VAN WIJK J.: Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2310–2316. 8

[DAW13] DANG T. N., ANAND A., WILKINSON L.: Timeseer: Scagnostics for high-dimensional time series. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 470–483. 7

[DBW*12] DEMIR D., BEKETAYEV K., WEBER G. H., BREMER P.-T., PASCUCCI V., HAMANN. B.: Topology exploration with hierarchical landscapes. In *Proceedings of the Workshop at SIGGRAPH Asia* (2012), pp. 147–154. 9, 13

[DCK*12] DUKE D., CARR H., KNOLL A., SCHUNCK N., NAM H. A., STASZCZAK A.: Visualizing nuclear scission through a multifield extension of topological analysis. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2033–2040. 6, 14

[DK10] DASGUPTA A., KOSARA R.: Pargnostics: Screen-space metrics for parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1017–1026. 11

[DKLP02] DJURCILOV S., KIM K., LERMUSIAUX P., PANG A.: Visualizing scalar volumetric data with uncertainty. *Computers and Graphics 26* (2002), 239–248. 14

[DOL03] DE OLIVEIRA M. C. F., LEVKOWITZ H.: From visual data exploration to visual data mining: A survey. *IEEE Transactions on Visualization and Computer Graphics 9*, 3 (2003), 378–394. 1, 11

[Don00] DONOHO D. L.: High-dimensional data analysis: The curses and blessings of dimensionality. AMS Lecture: Math Challenges of the 21st Century, 2000. 13

[dSMVJ09] DE SILVA V., MOROZOV D., VEJDEMO-JOHANSSON M.: Persistent cohomology and circular coordinates. In *Proceedings 25th Annual Symposium on Computational Geometry* (2009), pp. 227–236. 6

[DST04] DE SILVA V., TENENBAUM J. B.: *Sparse multidimensional scaling using landmark points*. Tech. rep., Technical report, Stanford University, 2004. 4, 12

[DW13] DEMIR I., WESTERMANN R.: Progressive high-quality response surfaces for visually guided sensitivity analysis. *Computer Graphics Forum 32*, 3pt1 (2013), 21–30. 5

[DWA10] DANG T. N., WILKINSON L., ANAND A.: Stacking graphic elements to avoid over-plotting. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1044–1052. 7

[ED07] ELLIS G., DIX A.: A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1216–1223. 1

[EDF08] ELMQVIST N., DRAGICEVIC P., FEKETE J.-D.: Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1539–1148. 10

[EH08] EDELSBRUNNER H., HARER J.: Persistent homology – a survey. *Contemporary Mathematics 453* (2008), 257. 6

[EH10] EDELSBRUNNER H., HARER J.: *Computational Topology - an Introduction.* American Mathematical Society, 2010. 6

[EHNP03] EDELSBRUNNER H., HARER J., NATARAJAN V., PASCUCCI V.: Morse-Smale complexes for piece-wise linear 3-manifolds. In *Proceedings 19th Annual symposium on Computational geometry* (2003), pp. 361–370. 6

[EHZ03] EDELSBRUNNER H., HARER J., ZOMORODIAN A. J.: Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Discrete and Computational Geometry 30*, 87-107 (2003). 6

[ERHH11] ENGEL D., ROSENBAUM R., HAMANN B., HAGEN H.: Structural decomposition trees. *Computer Graphics Forum 30*, 3 (2011), 921–930. 9

[EST07] ELMQVIST N., STASKO J., TSIGAS P.: Datameadow: A visual canvas for analysis of large-scale multivariate data. In *IEEE Symposium on Visual Analytics Science and Technology* (2007), pp. 187–194. 8

[FBT*10] FERDOSI B. J., BUDDELMEIJER H., TRAGER S., WILKINSON M. H., ROERDINK J. B.: Finding and visualizing relevant subspaces for clustering high-dimensional astronomical data using connected morphological operators. In *IEEE Symposium on Visual Analytics Science and Technology* (2010), IEEE, pp. 35–42. 5, 7, 11, 13

[FCI05] FANEA E., CARPENDALE S., ISENBERG T.: An interactive 3d integration of parallel coordinates and star glyphs. In *IEEE Symposium on Information Visualization* (2005), pp. 149–156. 8

[FR11] FERDOSI B. J., ROERDINK J. B.: Visualizing high-dimensional structures by dimension ordering and filtering using subspace analysis. *Computer Graphics Forum 30*, 3 (2011), 1121–1130. 7

[FT74] FRIEDMAN J., TUKEY J.: A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers C-23*, 9 (1974), 881–890. 5, 13

[FWR00] FUA Y.-H., WARD M., RUNDENSTEINER E.: Structure-based brushes: a mechanism for navigating hierarchically organized data and information spaces. *IEEE Transactions on Visualization and Computer Graphics 6*, 2 (2000), 150–159. 9

[GBPH08] GYULASSY A., BREMER P.-T., PASCUCCI V., HAMANN B.: A practical approach to Morse-Smale complex computation: Scalability and generality. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1619–1626. 6

[GBPW10] GERBER S., BREMER P., PASCUCCI V., WHITAKER R.: Visual exploration of high dimensional scalar functions. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1271–1280. 3, 5, 6, 13, 14

[GGL*14] GRATZL S., GEHLENBORG N., LEX A., PFISTER H., STREIT M.: Domino: Extracting, comparing, and manipulating subsets across multiple tabular datasets. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2023–2032. 12

[Ghr08] GHRIST R.: Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society 45* (2008), 61–75. 6

[Gle13] GLEICHER M.: Explainers: Expert explorations with crafted projections. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2042–2051. 4, 12, 14

[GLG*13] GRATZL S., LEX A., GEHLENBORG N., PFISTER H., STREIT M.: Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2277–2286. 12

[GNP*05] GYULASSY A., NATARAJAN V., PASCUCCI V., BREMER P.-T., HAMANN B.: Topology-based simplification for feature extraction from 3D scalar fields. In *Proceedings of IEEE Visualization* (2005), pp. 535–542. 6

[GPL*11] GENG Z., PENG Z., LARAMEE R., ROBERTS J., WALKER R.: Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2572–2580. 8

[Guo03] GUO D.: Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. *Information Visualization 2*, 4 (2003), 232–246. 7

[GWRR11] GUO Z., WARD M., RUNDENSTEINER E., RUIZ C.: Pointwise local pattern exploration for sensitivity analysis. In *IEEE Conference on Visual Analytics Science and Technology* (2011), pp. 131–140. 8

[GXWY10] GUO P., XIAO H., WANG Z., YUAN X.: Interactive local clustering operations for high dimensional data in parallel coordinates. In *IEEE Pacific Visualization Symposium* (2010), pp. 97–104. 12

[GXY12] GUO H., XIAO H., YUAN X.: Scalable multivariate volume visualization and analysis based on dimension projection and parallel coordinates. *IEEE transactions on visualization and computer graphics* (2012). 13

[HBM*13] HU X., BRADEL L., MAITI D., HOUSE L., NORTH C.: Semantics of directly manipulating spatializations. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2052–2059. 12, 14

[HDK*07] HAO M., DAYAL U., KEIM D., MORENT D., SCHNEIDEWIND J.: Intelligent visual analytics queries. In *IEEE Symposium on Visual Analytics Science and Technology* (2007), pp. 91–98. 11

[HG02] HOFFMAN P. E., GRINSTEIN G. G.: A survey of visualizations for high-dimensional data mining. *Information visualization in data mining and knowledge discovery* (2002), 47–82. 1

[HGM*97] HOFFMAN P., GRINSTEIN G., MARX K., GROSSE I., STANLEY E.: DNA visual and analytic data mining. In *Proceedings of IEEE Visualization* (1997), pp. 437–441. 7

[HO10] HURLEY C., OLDFORD R.: Pairwise display of high-dimensional information via eulerian tours and hamiltonian decompositions. *Journal of Computational and Graphical Statistics 19*, 4 (2010). 7

[HR07] HEER J., ROBERTSON G. G.: Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1240–1247. 10

[HS89] HASTIE T., STUETZLE W.: Principal curves. *Journal of the American Statistical Association 84*, 406 (1989), 502–516. 5

[HSKIH07] HAGH-SHENAS H., KIM S., INTERRANTE V., HEALEY C.: Weaving versus blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1270–1277. 11

[HTE11] HURTER C., TELEA A., ERSOY O.: Moleview: An attribute and structure-based semantic lens for large element-based plots. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2600–2609. 10, 12

[HVW10] HOLTEN D., VAN WIJK J. J.: Evaluation of cluster identification performance for different PCP variants. *Computer Graphics Forum 29*, 3 (2010), 793–802. 10

[HW09] HEINRICH J., WEISKOPF D.: Continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1531–1538. 10

[HW10] HARVEY W., WANG Y.: Topological landscape ensembles for visualization of scalar-valued functions. *Computer Graphics Forum 29*, 3 (2010), 993–1002. 9, 13

[HW13] HEINRICH J., WEISKOPF D.: State of the art of parallel coordinates. *STAR Proceedings of Eurographics 2013* (2013), 95–116. 1

[ID91] INSELBERG A., DIMSDALE B.: Parallel coordinates. In *Human-Machine Interactive Systems*. Springer, 1991, pp. 199–233. 7

[IML13] IM J.-F., MCGUFFIN M., LEUNG R.: Gplom: The generalized plot matrix for visualizing multidimensional multivariate data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2606–2614. 3

[Ins09] INSELBERG A.: *Parallel Coordinates : Visual Multidimensional Geometry and its Applications*. Springer, 2009. 1, 7

[JC08] JOHANSSON J., COOPER M.: A screen space quality method for data abstraction. *Computer Graphics Forum 27*, 3 (2008), 1039–1046. 11, 12

[JJ09] JOHANSSON S., JOHANSSON J.: Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 993–1000. 7

[JLJC05] JOHANSSON J., LJUNG P., JERN M., COOPER M.: Revealing structure within clustered parallel coordinates displays. In *IEEE Symposium on Information Visualization* (2005), IEEE, pp. 125–132. 10, 12

[JN02] JAYARAMAN S., NORTH C.: A radial focus+context visualization for multi-dimensional functions. In *Proceedings of IEEE Visualization* (2002), pp. 443–450. 8

[Joh04] JOHNSON C. R.: Top scientific visualization research problems. *IEEE Computer Graphics and Applications* (2004). 14

[Jol05] JOLLIFFE I.: *Principal component analysis*. Wiley Online Library, 2005. 3

[JPC*11] JOIA P., PAULOVICH F., COIMBRA D., CUMINATO J., NONATO L.: Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2563–2571. 4, 12

[JZF*09] JEONG D. H., ZIEMKIEWICZ C., FISHER B., RIBARSKY W., CHANG R.: iPCA: An interactive system for pca-based visual analytics. *Computer Graphics Forum 28*, 3 (2009), 767–774. 3

[Kan00] KANDOGAN E.: Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *IEEE Information Visualization Symposium, Late Breaking Hot Topics* (2000), pp. 9–12. 7

[KC03] KOREN Y., CARMEL L.: Visualization of labeled data using linear transformations. In *IEEE Symposium on Information Visualization* (2003), pp. 121–128. 4, 11

[Kei02] KEIM D. A.: Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (2002), 1–8. 1, 11

[KGZ*12] KUHNE L., GIESEN J., ZHANG Z., HA S., MUELLER K.: A data-driven approach to hue-preserving color-blending. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2122–2129. 11

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 495–513. 1, 3, 12

[KHL*01] KEIM D., HAO M., LADISCH J., HSU M., DAYAL U.: Pixel bar charts: a new technique for visualizing large multi-attribute data sets without aggregation. In *IEEE Symposium on Information Visualization* (2001), pp. 113–120. 9, 11

[KK94] KEIM D., KRIEGEL H.-P.: Visdb: database exploration using multidimensional visualization. *Computer Graphics and Applications, IEEE 14*, 5 (1994), 40–49. 9

[KK96] KEIM D. A., KRIEGEL H.-P.: Visualization techniques for mining large databases: A comparison. *Knowledge and Data Engineering, IEEE Transactions on 8*, 6 (1996), 923–938. 11

[KKA95] KEIM D., KRIEGEL H.-P., ANKERST M.: Recursive pattern: a technique for visualizing very large amounts of data. In *Proceedings of IEEE Visualization* (1995), pp. 279–286, 463. 9

[Kru64] KRUSKAL J. B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika 29*, 1 (1964), 1–27. 4

[KS02] KREUSELER M., SCHUMANN H.: A flexible approach for visual data mining. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (2002), 39–51. 9

[LA94] LEUNG Y. K., APPERLEY M. D.: A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction 1*, 2 (1994), 126–160. 10, 12

[LAE*12] LEHMANN D. J., ALBUQUERQUE G., EISEMANN M., MAGNOR M., THEISEL H.: Selecting coherent and relevant plots in large scatterplot matrices. *Computer Graphics Forum 31*, 6 (2012), 1895–1908. 7

[LAK*11] LAWRENCE J., ARIETTA S., KAZHDAN M., LEPAGE

D., O'Hagan C.: A user-assisted approach to visualizing multidimensional images. *IEEE Transactions on Visualization and Computer Graphics 17*, 10 (2011), 1487–1498. 3

[LMZ*14] Lee J. H., McDonnell K. T., Zelenyuk A., Imre D., Mueller K.: A structure-based distance metric for high-dimensional space exploration with multidimensional scaling. *IEEE Transations on Visualization and Computer Graphics 20*, 3 (2014), 351–364. 4

[LPM03] Lee A. B., Pedersen K. S., Mumford D.: The nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision 54*, 1-3 (2003), 83–103. 6

[LT11] Lehmann D., Theisel H.: Features in continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1912–1921. 10

[LT13] Lehmann D. J., Theisel H.: Orthographic star coordinates. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2615–2624. 7

[LV09] Lee J. A., Verleysen M.: Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing 72*, 7 (2009), 1431–1443. 4

[LWBP14] Liu S., Wang B., Bremer P.-T., Pascucci V.: Distortion-guided structure-driven interactive exploration of high-dimensional data. *Computer Graphics Forum 33*, 3 (2014), 101–110. 4, 12, 14

[LWT*14a] Liu S., Wang B., Thiagarajan J. J., Bremer P.-T., Pascucci V.: Multivariate volume visualization through dynamic projections. *Large Data Analysis and Visualization (LDAV), 2014 IEEE Symposium on* (2014). 13, 14

[LWT*14b] Liu S., Wang B., Thiagarajan J. J., Bremer P.-T., V V. P.: *Visual exploration of high-dimensional data: Subspace analysis through dynamic projections*. Tech. Rep. UUSCI-2014-003, SCI Institute, University of Utah, 2014. 14

[MG13] Mayorga A., Gleicher M.: Splatterplots: Overcoming overdraw in scatter plots. *IEEE Transactions on Visualization and Computer Graphics 19*, 9 (2013), 1526–1538. 10, 13

[MLGH13] Mokbel B., Lueks W., Gisbrecht A., Hammer B.: Visualizing the quality of dimensionality reduction. *Neurocomputing 112* (2013), 109–123. 4

[MM08] McDonnell K. T., Mueller K.: Illustrative parallel coordinates. *Computer Graphics Forum 27*, 3 (2008), 1031–1038. 10, 11

[MPG*14] Muhlbacher T., Piringer H., Gratzl S., Sedlmair M., Streit M.: Opening the black box: Strategies for increased user involvement in existing algorithm implementations. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1643–1652. 11

[MRC02] Morrison A., Ross G., Chalmers M.: A hybrid layout algorithm for sub-quadratic multidimensional scaling. In *IEEE Symposium on Information Visualization* (2002), IEEE, pp. 152–158. 11

[Mun14] Munzner T.: *Visualization Analysis and Design*. CRC Press, 2014. 1, 2

[MW10] Migut M., Worring M.: Visual exploration of classification models for risk assessment. In *IEEE Symposium on Visual Analytics Science and Technology* (2010), pp. 11–18. 13

[NH06] Novotny M., Hauser H.: Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 893–900. 7

[NLC11] Nicolau M., Levine A. J., Carlsson G.: Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. In *Proceedings of the National Academy of Sciences* (2011), vol. 108, pp. 7265–7270. 6

[NM13] Nam J. E., Mueller K.: Tripadvisor-nd: A tourism-inspired high-dimensional space exploration framework with overview and detail. *IEEE Transactions on Visualization and Computer Graphics 19*, 2 (2013), 291–305. 5, 10

[OHJ*11] Oesterling P., Heine C., Janicke H., Scheuermann G., Heyer G.: Visualization of high-dimensional point clouds using their density distribution's topology. *IEEE Transactions on Visualization and Computer Graphics 17*, 11 (2011), 1547–1559. 9

[OHJS10] Oesterling P., Heine C., Jänicke H., Scheuermann G.: Visual analysis of high dimensional point clouds using topological landscape. In *IEEE Pacific Visualization Symposium* (2010), pp. 113–120. 9

[OHWS13] Oesterling P., Heine C., Weber G., Scheuermann G.: Visualizing nd point clouds as topological landscape profiles to guide local data analysis. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 514–526. 9

[PBK10] Piringer H., Berger W., Krasser J.: Hypermoval: Interactive visual validation of regression models for real-time simulation. In *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization* (2010), EuroVis'10, Eurographics Association, pp. 983–992. 3, 5

[PCMS09] Pascucci V., Cole-McLaughlin K., Scorzelli G.: The toporrery: Computation and presentation of multi-resolution topology. *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration* (2009), 19–40. 9

[PdSABD*12] Portes dos Santos Amorim E., Brazil E. V., Daniels J., Joia P., Nonato L. G., Sousa M. C.: ilamp: Exploring high-dimensional spacing through backward multidimensional projection. In *IEEE Conference on Visual Analytics Science and Technology* (2012), IEEE, pp. 53–62. 12, 14

[PEP*11a] Paulovich F., Eler D., Poco J., Botha C., Minghim R., Nonato L.: Piece wise laplacian-based projection for interactive data exploration and organization. *Computer Graphics Forum 30*, 3 (2011), 1091–1100. 4, 12

[PEP*11b] Poco J., Etemadpour R., Paulovich F., Long T., Rosenthal P., Oliveira M., Linsen L., Minghim R.: A framework for exploring multidimensional data with 3d projections. *Computer Graphics Forum 30*, 3 (2011), 1111–1120. 12

[PNML08] Paulovich F., Nonato L., Minghim R., Levkowitz H.: Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics 14*, 3 (2008), 564–575. 4, 12

[PSBM07] Pascucci V., Scorzelli G., Bremer P.-T., Mascarenhas A.: Robust on-line computation of reeb graphs: Simplicity and speed. *ACM Transactions on Graphics 26*, 3 (2007). 6

[PSN10] Paulovich F., Silva C., Nonato L.: Two-phase mapping for projecting massive data sets. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1281–1290. 4, 12

[PWB*09] Potter K., Wilson A., Bremer P.-T., Williams D., Pascucci V., Johnson C.: Ensemblevis: A flexible approach for the statistical visualization of ensemble data. In *Proceedings of IEEE Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts* (2009). 14

[PWR04] PENG W., WARD M. O., RUNDENSTEINER E. A.: Clutter reduction in multi-dimensional data visualization using dimension reordering. In *IEEE Symposium on Information Visualization* (2004), IEEE, pp. 89–96. 7, 12

[RC94] RAO R., CARD S. K.: The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1994), ACM, pp. 318–322. 10

[Rd00] RHEINGANS P., DESJARDINS M.: Visualizing high-dimensional predictive model quality. In *Proceedings of IEEE Visualization* (2000), pp. 493–496. 13

[Ree46] REEB G.: 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. Science Paris 222* (1946), 847–849. 6

[RPH08] REDDY C. K., POKHARKAR S., HO T. K.: Generating hypotheses of trends in high-dimensional data skeletons. In *IEEE Symposium on Visual Analytics Science and Technology* (2008), IEEE, pp. 139–146. 5

[RRB*04] ROSARIO G. E., RUNDENSTEINER E. A., BROWN D. C., WARD M. O., HUANG S.: Mapping nominal values to numbers for effective visualization. *Information Visualization 3*, 2 (2004), 80–95. 3

[RS00] ROWEIS S. T., SAUL L. K.: Nonlinear dimensionality reduction by locally linear embedding. *Science 290*, 5500 (2000), 2323–2326. 4, 11

[RW06] RASMUSSEN C. E., WILLIAMS C. K. I.: *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006. 5

[RZH12] ROSENBAUM R., ZHI J., HAMANN B.: Progressive parallel coordinates. In *IEEE Pacific Visualization Symposium* (2012), pp. 25–32. 7

[Shn92] SHNEIDERMAN B.: Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG) 11*, 1 (1992), 92–99. 9

[SLBC03] SWAYNE D. F., LANG D. T., BUJA A., COOK D.: GGobi: evolving from XGobi into an extensible framework for interactive data visualization. *Computational Statistics & Data Analysis 43*, 4 (2003), 423–444. 9, 12

[Sma61] SMALE S.: On gradient dynamical systems. *The Annals of Mathematics 74* (1961), 199–206. 6

[SMC07] SINGH G., MEMOLI F., CARLSSON G.: Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Symposium on Point Based Graphics* (2007), pp. 91–100. 3, 6, 13

[SMT13] SEDLMAIR M., MUNZNER T., TORY M.: Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2634–2643. 10

[SNLH09] SIPS M., NEUBERT B., LEWIS J. P., HANRAHAN P.: Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum 28*, 3 (2009), 831–838. 7

[SS04a] SEO J., SHNEIDERMAN B.: A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. In *IEEE Symposium on Information Visualization* (2004), IEEE, pp. 65–72. 7, 13

[SS04b] SMOLA A. J., SCHÖLKOPF B.: A tutorial on support vector regression. *Statistics and Computing 14*, 3 (2004), 199–222. 5

[SS06] SEO J., SHNEIDERMAN B.: Knowledge discovery in high-dimensional data: Case studies and a user survey for the rank-by-feature framework. *IEEE Transactions on Visualization and Computer Graphics 12*, 3 (2006), 311–322. 7

[SSK06] SCHNEIDEWIND J., SIPS M., KEIM D. A.: Pixnostics: Towards measuring the value of visualization. In *IEEE Symposium on Visual Analytics Science and Technology* (2006), IEEE, pp. 199–206. 11

[SSK10] SEIFERT C., SABOL V., KIENREICH W.: Stress maps: analysing local phenomena in dimensionality reduction based visualisations. In *IEEE International Symposium on Visual Analytics Science and Technology.* (2010). 4

[STH02] STOLTE C., TANG D., HANRAHAN P.: Polaris: a system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (2002), 52–65. 11

[STH03] STOLTE C., TANG D., HANRAHAN P.: Multiscale visualization using data cubes. *IEEE Transactions on Visualization and Computer Graphics 9*, 2 (2003), 176–187. 11

[SvLB10] SCHRECK T., VON LANDESBERGER T., BREMM S.: Techniques for precision-based visual analysis of projected data. *Information Visualization 9*, 3 (2010), 181–193. 4, 14

[SVW10] SHRINIVASAN Y. B., VAN WIJK J. J.: Supporting exploratory analysis with the select & slice table. *Computer Graphics Forum 29*, 3 (2010), 803–812. 12

[SW09] SANFTMANN H., WEISKOPF D.: Illuminated 3d scatterplots. *Computer Graphics Forum 28*, 3 (2009), 751–758. 10, 12

[SW12] SANFTMANN H., WEISKOPF D.: 3d scatterplot navigation. *IEEE Transactions on Visualization and Computer Graphics 18*, 11 (2012), 1969–1978. 12

[SZD*10] SANYAL J., ZHANG S., DYER J., MERCER A., AMBURN P., MOORHEAD R. J.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1421 – 1430. 14

[TAE*09] TATU A., ALBUQUERQUE G., EISEMANN M., SCHNEIDEWIND J., THEISEL H., MAGNOR M., KEIM D.: Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *IEEE Symposium on Visual Analytics Science and Technology* (2009), IEEE, pp. 59–66. 7, 11, 13

[TDSL00] TENENBAUM J. B., DE SILVA V., LANGFORD J. C.: A global geometric framework for nonlinear dimensionality reduction. *Science 290*, 5500 (2000), 2319–2323. 4

[TFA*11] TAM G. K. L., FANG H., AUBREY A. J., GRANT P. W., ROSIN P. L., MARSHALL D., CHEN M.: Visualization of time-series data in parameter space for understanding facial dynamics. *Computer Graphics Forum 30*, 3 (2011), 901–910. 3

[TFH11] TURKAY C., FILZMOSER P., HAUSER H.: Brushing dimensions-a dual visual analysis model for high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2591–2599. 4, 11, 12

[TFO09] TAKAHASHI S., FUJISHIRO I., OKADA M.: Applying manifold learning to plotting approximate contour trees. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1185–1192. 14

[TJHH14] TURKAY C., JEANQUARTIER F., HOLZINGER A., HAUSER H.: On computationally-enhanced visual analysis of heterogeneous data and its application in biomedical informatics. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Springer, 2014, pp. 117–140. 11

[TLLH12] TURKAY C., LUNDERVOLD A., LUNDERVOLD A. J., HAUSER H.: Representative factor generation for the interactive visual analysis of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2621–2630. 4, 11

[TM03] TEOH S. T., MA K.-L.: Paintingclass: interactive construction, visualization and exploration of decision trees. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (2003), ACM, pp. 667–672. 13

[TM05] TZENG F.-Y., MA K.-L.: Opening the black box - data driven visualization of neural networks. In *Proceedings of IEEE Visualization* (2005), pp. 383–390. 13

[TMF*12] TATU A., MAAS F., FARBER I., BERTINI E., SCHRECK T., SEIDL T., KEIM D.: Subspace search and visualization to make sense of alternative clusterings in high-dimensional data. In *IEEE Conference on Visual Analytics Science and Technology* (2012), IEEE, pp. 63–72. 5, 11, 12, 13

[TWSM*11] TORSNEY-WEIR T., SAAD A., MOLLER T., HEGE H.-C., WEBER B., VERBAVATZ J., BERGNER S.: Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1892–1901. 5

[vdEvW11] VAN DEN ELZEN S., VAN WIJK J.: Baobabview: Interactive construction and analysis of decision trees. In *IEEE Conference on Visual Analytics Science and Technology* (2011), pp. 151–160. 13

[Vid11] VIDAL R.: A tutorial on subspace clustering. *IEEE Signal Processing Magazine* (2011). 5, 13

[WAG05] WILKINSON L., ANAND A., GROSSMAN R.: Graph-theoretic scagnostics. In *IEEE Symposium on Information Visualization* (2005), vol. 0, p. 21. 7, 11, 12

[WAG06] WILKINSON L., ANAND A., GROSSMAN R.: High-dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions. *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1363–1372. 7, 13

[War94] WARD M. O.: Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of IEEE Visualization* (1994), pp. 326–333. 3

[War08] WARD M. O.: Multivariate data glyphs: Principles and practice. In *Handbook of Data Visualization*. Springer, 2008, pp. 179–198. 8

[Wat05] WATTENBERG M.: A note on space-filling visualizations and space-filling curves. In *IEEE Symposium on Information Visualization* (2005), pp. 181–186. 9

[WB94] WONG P. C., BERGERON R. D.: 30 years of multidimensional multivariate visualization. In *Proceedings of Scientific Visualization, Overviews, Methodologies, and Techniques* (1994), pp. 3–33. 1

[WBP07] WEBER G., BREMER P.-T., PASCUCCI V.: Topological landscapes: A terrain metaphor for scientific data. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1416–1423. 9

[WBP12] WEBER G. H., BREMER P.-T., PASCUCCI V.: Topological cacti: Visualizing contour-based statistics. *Topological Methods in Data Analysis and Visualization II Mathematics and Visualization* (2012), 63–76. 9

[Wea09] WEAVER C.: Conjunctive visual forms. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 929–936. 3

[WM04] WILLIAMS M., MUNZNER T.: Steerable, progressive multidimensional scaling. In *IEEE Symposium on Information Visualization* (2004), pp. 57–64. 4, 11, 12

[WO11] WADDELL A., OLDFORD R. W.: RnavGraph: A visualization tool for navigating through high-dimensional data. 10

[WPWR03] WANG J., PENG W., WARD M. O., RUNDENSTEINER E. A.: Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *IEEE Symposium on Information Visualization* (2003), IEEE, pp. 105–112. 9

[WSPVJ11] WANG B., SUMMA B., PASCUCCI V., VEJDEMO-JOHANSSON M.: Branching and circular features in high dimensional data. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1902–1911. 6, 14

[YGX*09] YUAN X., GUO P., XIAO H., ZHOU H., QU H.: Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1001–1008. 8, 10, 12

[YHW*07] YANG J., HUBBALL D., WARD M. O., RUNDENSTEINER E. A., RIBARSKY W.: Value and relation display: interactive visual exploration of large data sets with hundreds of dimensions. *IEEE Transactions on Visualization and Computer Graphics 13*, 3 (2007), 494–507. 9

[YPS*04] YANG J., PATRO A., SHIPING H., MEHTA N., WARD M., RUNDENSTEINER E.: Value and relation display for interactive exploration of high dimensional datasets. In *IEEE Symposium on Information Visualization* (2004), pp. 73–80. 9

[YRWG13] YUAN X., REN D., WANG Z., GUO C.: Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2625–2633. 4, 11

[YWR02] YANG J., WARD M. O., RUNDENSTEINER E. A.: Interring: An interactive tool for visually navigating and manipulating hierarchical structures. In *IEEE Symposium on Information Visualization* (2002), IEEE, pp. 77–84. 9

[ZCQ*09] ZHOU H., CUI W., QU H., WU Y., YUAN X., ZHUO W.: Splatting the lines in parallel coordinates. *Computer Graphics Forum 28*, 3 (2009), 759–766. 11, 13

[ZJGK10] ZIEGLER H., JENNY M., GRUSE T., KEIM D.: Visual market sector analysis for financial time series data. In *IEEE Symposium on Visual Analytics Science and Technology* (2010), pp. 83–90. 3

[Zom05] ZOMORODIAN A. J.: *Topology for Computing (Cambridge Monographs on Applied and Computational Mathematics)*. Cambridge University Press, 2005. 6

[ZSWR06] ZAIXIAN X., SHIPING H., WARD M., RUNDENSTEINER E.: Exploratory visualization of multivariate data with variable quality. In *IEEE Symposium on Visual Analytics Science and Technology* (2006), pp. 183–190. 14

Brief Biographies of the Authors

Shusen Liu received his bachelor degree in Biomedical Engineering and Computer Science from Huazhong University of Science and Technology, China, in 2009, where he worked at Wuhan National Laboratory for Optoelectronic on GPU accelerated biophotonics applications. Currently he is a PhD student at University of Utah. His research interests lie primarily in high-dimensional data visualization and multivariate volume visualization.

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

Bei Wang received her Ph.D. in Computer Science from Duke University in 2010. She is currently a Research Scientist at the Scientific Computing and Imaging Institute, University of Utah. Her main research interests are computational topology, computational geometry, scientific data analysis and visualization. She is also interested in computational biology and bioinformatics, machine learning and data mining. She is a member of the IEEE Computer Society.

Peer-Timo Bremer is a member of technical staff and project leader at the Center for Applied Scientific Computing (CASC) at the Lawrence Livermore National Laboratory (LLNL) and Associated Director for Research at the Center for Extreme Data Management, Analysis, and Visualization at the University of Utah. His research interests include large scale data analysis, performance analysis and visualization and he recently co-organized a Dagstuhl Perspectives workshop on integrating performance analysis and visualization. Prior to his tenure at CASC, he was a postdoctoral research associate at the University of Illinois, Urbana-Champaign. Peer-Timo earned a Ph.D. in Computer science at the University of California, Davis in 2004 and a Diploma in Mathematics and Computer Science from the Leipniz University in Hannover, Germany in 2000. He is a member of the IEEE Computer Society and ACM.

Valerio Pascucci is the founding Director of the Center for Extreme Data Management Analysis and Visualization (CEDMAV) of the University of Utah. Valerio 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. Previously, Valerio was a Group Leader and Project Leader in the Center for Applied Scientific Computing at the Lawrence Livermore National Laboratory, and Adjunct Professor of Computer Science at the University of California Davis. Prior to his CASC tenure, he was a senior research associate at the University of Texas at Austin, Center for Computational Visualization, CS and TICAM Departments. Valerio earned a Ph.D. in computer science at Purdue University in May 2000, and a EE Laurea (Master), at the University "La Sapienza" in Roma, Italy, in December 1993, as a member of the Geometric Computing Group. His recent research interest is in developing new methods for massive data analysis and visualization.