# Interactive Analysis of High-Dimensional Data using Visualization

Helwig Hauser Robert Kosara
VRVis Research Center in Vienna, Austria,
`http://www.VRVis.at/vis/`,
{`Hauser`,`Kosara`}`@VRVis.at`

## Overview

Visually displaying data with two or three dimensions is very common. Humans can easily recognize structures in the data (such as correlations in a scatterplot, trends in a line chart, etc.) and get a better impression of the data from images than from reading numbers. Relatively recently, it has become possible to visualize high-dimensional data, and to do so with the help of the computer. This not only makes it possible to quickly draw complex graphics, but also to interact with them. Interaction is key to visually analyzing high-dimensional data, and to finding complex relationships in them.

Information visualization (InfoVis) provides methods for the interactive exploration and analysis of high-dimensional data such as, results from complex numerical simulations, multi-dimensional product databases, etc. In this paper, a few of the central concepts of InfoVis are introduced: (1) *visualization with multiple views*, which often are (but not necessarily need to be) of different visualization types and which are *visually linked* to each other, especially when used in conjunction with *interactive brushing* (linking and brushing, L&B); (2) *focus-plus-context visualization* (F+C visualization) as a means to jointly support *zooming* into the visual depiction of the data while at the same time maintaining the *visual orientation* of the visualization user to support navigation in the visualization; and (3) the *potential combination* of visualization methods and such from statistics as an interesting perspective for future work.

## 1 Visualization of High-Dimensional Data

In visualization research, many different visualization techniques have been developed, which are good for different investigation purposes. Well-known examples are scatterplots and histograms. In addition to these (rather historic) approaches, other new techniques have been proposed such as parallel coordinates [9], icon- or pixel-oriented techniques [17, 11], as well as many others – Kosara et al. give a useful overview about visualization techniques [14].

Visualization exploits the powerful human visual system to effectively transport information from the outside world to the human apparatus of perception, recognition, cognition, and reasoning [20]. Because the effectiveness of visualization methods cannot be established simply by formal means, they are often tested for their effectiveness in empirical studies [15].

## 2 Multiple Views, Linking and Brushing

The use of multiple views [1] is one of the central paradigms in InfoVis. Displaying the data in several views makes it possible to communicate more information without overloading a single view with too much information. The user can see the information in each view separately, and can understand the connection between the views using interaction – because they are linked.

Visualization of this kind always follows the same principle: the same data is shown in several separate views (or components thereof). Each view usually shows another aspect of the data, either through the use of an alternative visualization technique, or

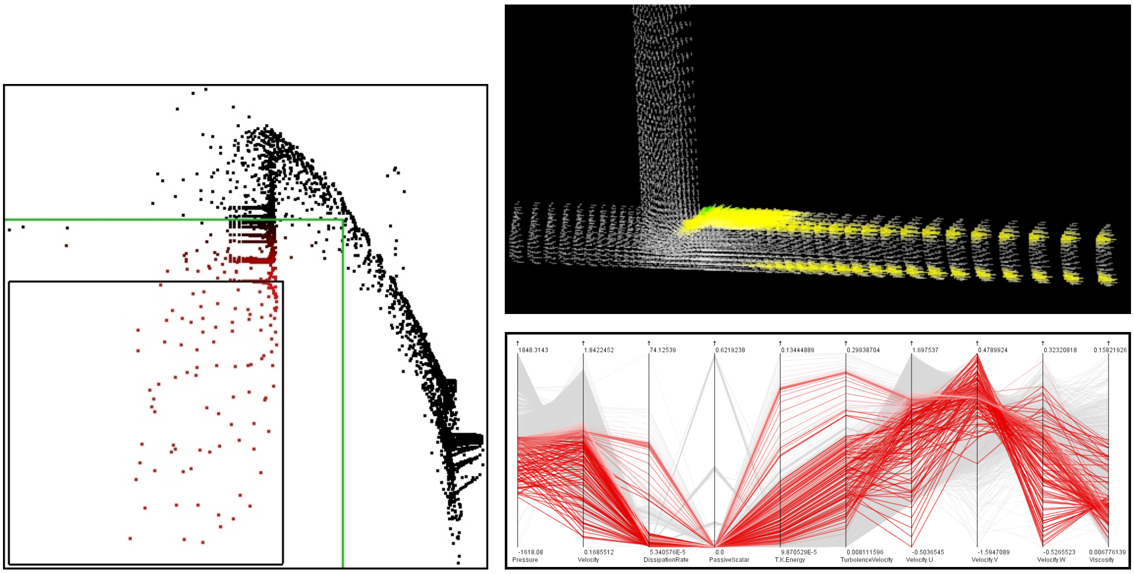

Figure 1: Linking and brushing, a sample visualization of a high-dimensional simulation dataset: in a scatter-plot (shown on the left side, two data dimensions), smooth brushing [4] was used to mark data-points of low pressure and low velocity; a linked 3D view (on the top right, spatial view) shows the same data with the brushed data-points high-lighted; thirdly, the parallel coordinates view (on the lower right, ten of the data dimensions shown) also shows the same data, also high-lighting the brushed sub-set.

through the use of a specific projection. For example, a dataset could be simultaneously visualized by the use of one scatterplot and a parallel coordinates view (different techniques) or by two scatterplots which show different dimensions of the dataset each (different projections).

To exploit the potential of data visualization with multiple views, it is essential that visualization cues which represent identical parts of the data in different views can be easily associated with each other visually. An often used solution to visually link separate views of one dataset is to choose the same color for visualization components which represent the same data items [19]. This visual *linking* between views becomes especially useful, when interactive *brushing* is supported in at least one of the views [2, 16]. Brushing means that the user can interactively select certain subsets of the data in one of the visualization views and at the same time study the high-dimensional characteristics of the selected data items in other (linked) views (the selected data items are all colored red, for example, and therefore stand out in the different visualization views simultaneously). If brushes can be applied, moved, altered, added/removed, and logically combined interactively, powerful analysis of high-dimensional data is possible through the means of interaction. To improve the quality of interactive work, we proposed advanced brushing techniques, e.g., smooth brushing [4] and angular brushing [8].

## 3 Interactive Focus+Context Visualization

One problem with the visualization of large datasets is that either an overview of data without details is conveyed, or the visualization has zoomed in onto specific details of the data without providing sufficient information about the conetxt of the depicted data. To overcome this problem, various techniques for *focus-plus-context* visualization (F+C visualization [14]) have been developed, with the goal of integrating both options of visualization: overview and details. Usually, spatial distortions are used to open up more space for the depiction of details in a visualization while still using the rest of the available space to show the rest of the data as context (in reduced form). The most prominent examples for distortion-oriented F+C visualization techniques are fisheye views [6, 10] and the document lens [18].

In recent work, we have demonstrated that focus+context visualization can be generalized to other visualization dimensions, as well [7]. Through the uneven use of graphics resources such as space, color, opacity, etc., a differentiated view can be

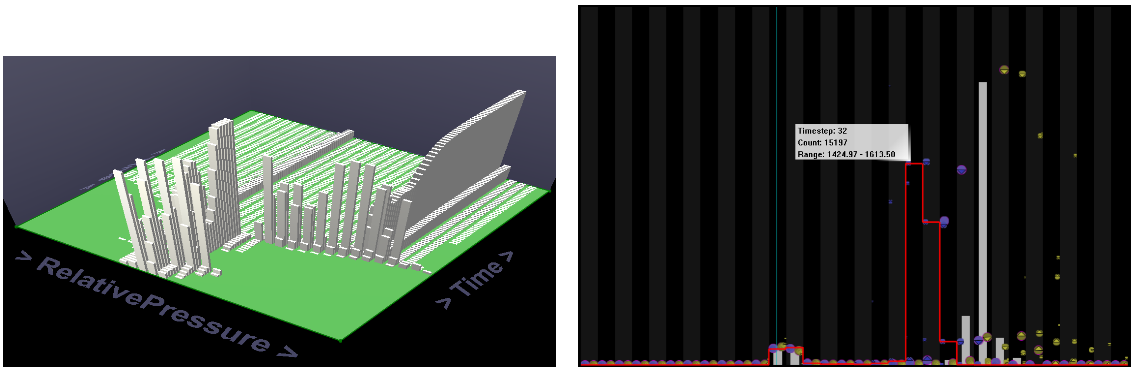


Figure 2: Histograms for time-dependent data. Left: TimeHistograms in 3D. One axis represents time, the other the relative pressure in the cells. The height of the bars shows the number of points in each bin. Right: TimeHistograms in 2D. In addition to the histogram for one time step (grey), the histograms for preceeding and following time steps are displayed with little yellow and blue disks. The user can point to a disk to see the histogram for that time step (red).

generated which locally focusses on specific, user-selected details whereas still providing additional overview of the data as context. Also, F+C visualization very well meets the approach of linking and brushing in multiple views – if certain data items are brushed in one view, F+C visualization can be used in the other views to visually differentiate between the selected data and all the rest.

# 4 The SimVis Application

Data in computational fluid dynamics (CFD) is usually large (hundred thousands data points) and high dimensional (15-25 dimensions). Simulations are also often done for processes that change over time, adding time as an additional dimenion (with 10 to 100 time steps for a data set).

We have developed a visualization system called SimVis [3, 5] which is capable of different visualization techniques (scatterplots, histograms, parallel coordinates, spatial 3D views, etc.) for CFD data. All these views are linked, and have been enhanced to work with time-dependent data. See Fig. 1 for a sample setup of linked SimVis views with a scatterplot, parallel coordinates and a spatial view.

Time plays quite a different role in this system than the other dimensions, because of its prime importance for physical processes. Histograms for time-dependent data (Fig. 2 [13]) provide the user with an overview of how the data changes, including direct comparisons of different time steps. But the display is not limited to showing the data, the user can also brush data in histograms, like in any other view, and this way get more insights into it.

# 5 Visualization and Statistics: Visual Data Mining

From the InfoVis point of view, the combination of visualization techniques with solutions from statistics and data mining seems very promising. The potential of this combination (called *visual data mining* [12]) arises from the fact that InfoVis and statistics pursue different approaches to reach the same goal: provide the user with insight into complex datasets. Mixing visualization with traditional data mining and other tools provides more possibilities for the user to take part in the process and to add information to the analysis.

Visualization can serve both as a tool for communicating the results of mathematical data mining and as a tool for data analysis itself. Visualization can also be supported with the results of statistical analysis to improve the display or interaction [21]. The combination of visualization with statistical analysis provides more and faster insight into data, as well as easier communication of results.

# Acknowledgements

Parts of this work have been done at the VRVis Research Center (`http://www.VRVis.at/`) in Vienna, Austria, which is funded by an Austrian governmental research program called Kplus.

## References


[1] M. Baldano, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proc. of Advanced Visual Interfaces*, pages 110–119, 2000.

[2] R. Becker and W. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[3] H. Doleisch, M. Gasser, and H. Hauser. Interactive feature specification for focus+context visualization of complex simulation data. In *Proc. of the Joint IEEE TCVG – EG Symp. on Vis.*, pages 239–248, 2003.

[4] H. Doleisch and H. Hauser. Smooth brushing for focus+context visualization of simulation data in 3D. *Journal of WSCG*, 10(1):147–154, 2002.

[5] H. Doleisch, M. Mayer, M. Gasser, R. Wanker, and H. Hauser. Case study: Visual analysis of complex, time-dependent simulation results of a diesel exhaust system. In *Proc. of the Joint IEEE TCVG – EG Symposium on Visualization*, 2004.

[6] G. Furnas. Generalized fisheye views. In *Proc. of the ACM CHI'86 Conference on Human Factors in Computing Systems*, pages 16–23, 1986.

[7] H. Hauser. Generalizing focus+context visualization. In *Proc. of the Dagstuhl 2003 Seminar on Scientific Visualization*. 2003. to appear (VRVis Tecchnical Report TR-VRVis-2003-037).

[8] H. Hauser, F. Ledermann, and H. Doleisch. Angular brushing for extended parallel coordinates. In *Proc. of the IEEE Symposium on Information Visualization*, pages 127–130, 2002.

[9] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multidimensional geometry. In *Proc. of IEEE Visualization '90*, pages 361–378, 1990.

[10] T. Keahey and E. Robertson. Nonlinear magnification fields. In *Proc. of the IEEE Symp. on Information Visualization*, pages 51–58, 1997.

[11] D. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):59–78, 2000.

[12] D. Keim, W. Müller, and H. Schumann. Visual data mining. In EUROGRAPHICS *2002 State-of-the-Art Reports*, pages 49–68, 2002.

[13] R. Kosara, F. Bendix, and H. Hauser. Time-histograms for large, time-dependent data. In *Joint Eurographics – IEEE TCVG Symposium on Visualization*, 2004.

[14] R. Kosara, H. Hauser, and D. Gresh. An interaction view on information visualization. In EUROGRAPHICS *2003 State-of-the-Art Reports*, pages 123–137, 2003.

[15] R. Kosara, C. G. Healey, V. Interrante, D. H. Laidlaw, and C. Ware. Thoughts on user studies: Why, how, and when. *IEEE Computer Graphics & Applications (CG&A), Visualization Viewpoints*, 23(4), July/August 2003.

[16] A. Martin and M. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Proc. of IEEE Visualization '95*, pages 271–278, 1995.

[17] R. Pickett and G. Grinstein. Iconographic displays for visualizing multidimensional data. In *Proc. of the IEEE Conf. on Systems, Man and Cybernetics*, pages 514–519. IEEE Press, 1988.

[18] G. Robertson and J. Mackinlay. The document lens. In *Proc. of the ACM Symposium on User Interface Software and Technology*, pages 101–108, 1993.

[19] D. Swayne, D. Cook, and A. Buja. XGobi: Interactive dynamic data visualization in the X windows system. *Journal of Computational and Graphical Statistics*, 7(1):113–130, 1998.

[20] C. Ware. *Information Visualization – Perception for Design*. Morgan Kaufmann Publishers, 2000.

[21] J. Yang, M. Ward, E. Rundensteiner, and S. Huang. Visual hierarchical dimension reduction for exploration of high dimensional datasets. In *Proc. of the Joint IEEE TCVG – EG Symp. on Vis.*, pages 19–28, 2003.