Problem-oriented visualisation of multi-dimensional data sets

Susanne Lange
University of Rostock
Department of Computer Science
Institute for Computer Graphics
Albert-Einstein-Str. 21
18051 Rostock
phone: 0381 498 3418
fax: 0381 498 3426
email: slange@informatik.uni-rostock.de

Heidrun Schumann
University of Rostock
Department of Computer Science
Institute for Computer Graphics
Albert-Einstein-Str. 21
18051 Rostock
phone: 0381 498 3421
fax: 0381 498 3426
email: schumann@informatik.uni-rostock.de

Wolfgang Müller
Darmstadt Technical University
Department of Computer Science
Interactive Graphics Systems Group
Wilhelminenstr. 7
64283 Darmstadt
phone: 06151 155 136
fax: 06151 155 199
email:mueller@igd.fhg.de

Detlef Krömker
Fraunhofer IGD
Department Animation &
HD Image Communication
Wilhelminenstr. 7
64283 Darmstadt
phone: 06151 155 132
fax: 06151 155 199
email:kroemker@igd.fhg.de

Abstract

The amount of data produced in the fields of engineering and science by measurements and simulations is steadily increasing. Scientific Visualisation has proven to be often the only reasonable approach to analyse this data to gain knowledge about the underlying processes and relations. However, the decision for the right visualisation methods is not easy at all. In general, the user must be an expert to generate an effective visualisation which considers the goal in mind and renders the characteristics of the given data set without producing artefacts and hereby misleading visualisations. This problem is of great impact since such wrong visualisations may lead to wrong conclusions.

To support the user in the process of the generation of a visualisation it is necessary to provide guidance and tools which recommend visualisation methods and visualisation mappings, or even select an appropriate method. In this paper we discuss the various aspects of such visualisation support systems and present Vis-Wizz, a prototype of a problem-driven visualisation assistant.

Keywords
scientific visualisation, intelligent visualisation systems, visualisation support, visualisation assistant, multi-dimensional data, presentation quality

1 Introduction

Huge amounts of data are produced daily by measurement, observation, simulation and calculation. It is estimated that only a quarter of that data can be stored and only a quarter of the stored data can be analysed. The latter is mainly caused by the lack of effective tools and appropriate guidance for data analysis.

Scientific Visualisation represents one approach which is more and more applied in the problem area of analysing unknown data. Visualisation has proven to be a good and often the only reasonable approach to gain knowledge about the underlying processes and relations of the data. However, although a good variety of sophisticated visualisation techniques are provided so far by visualisation systems, visualisation did not get utilisation in the expected degree for several reasons:

- Most of the visualisation systems do not offer standardised interfaces for data import and export. The accepted formats are generally not those in which the data are created or held.
- The user interface design does not meet the terminology of the underlying application.
- For several data classes (particularly for multi-dimensional data) appropriate visualisation techniques are rarely supplied.
- Most important, there is no support of selection and use of different visualisation methods.

For special applications these problems can be avoided by providing a turnkey system. However, it seems not effective always to develop yet another visualisation system. There are general purpose visualisation systems addressing a variety of application fields and allowing to build ones own visualisation application. Since these systems (application builders) are powerful and expected to spread widely over the data producing fields, it should be aimed at providing assistance to its use and adjustment. Still, the above mentioned drawbacks fully apply to these systems.

The research described in this paper focuses on the enhancement of standard visualisation systems by an automatic generation of the visualisation network of an application builder. This involves the process of selecting a proper visualisation method and results in an automatic assembling of the application builder’s modules.
Starting from a classification of known approaches we describe a new approach to a problem-driven visualisation. Then we present our visualisation support tool and give some results.

2 Visualisation support - a classification

Modern visualisation systems provide a large amount of visualisation methods and techniques. However, the decision for the right visualisation methods and visualisation mapping is not easy at all. In general, the user must be an expert to generate an effective visualisation which considers the goal in mind and renders the characteristics of the given data set without producing artefacts and hereby misleading visualisations. Thus, there is a need for assistance tools which support the user during the modelling of the visualisation pipeline and allow even non-experienced users to visualise their data in an efficient way in an adequate amount of time. There are a number of different approaches to this problem. In this section we try to classify these approaches and to discuss their characteristics and their potential limitations.

During the modelling of a visualisation process a user may receive assistance in a problem-driven or a problem-independent way. Problem-independent support does not recognise the current context and the actual questions of the user when providing the assistance, though such systems are not limited to simple help mechanisms. Visualisation methodologies as proposed by Tufte [Tuft 83] and Bertin [Bert 83], and rule systems as introduced by Senay and Ignatius [SeFe 91] and others (see [ArLS 94] for a more detailed discussion) may represent the bases for more sophisticated solutions. An example of an interactive multimedia visualisation assistant based on this work is presented in [Koop 95].

More interesting is the idea of a problem-driven support of the user, where the system reacts on the current context and directly takes into account the various aspects which influence the effectiveness of a visualisation. The most important of these aspects are the data characteristics, which is not only the structure of the data but also the semantics, the visualisation goal ([Robe 90][Wehr 90]), the application domain, which affects for example the valid visualisation metaphors, the display characteristics, and, last not least, the visual capabilities of the observer. To consider these aspects in the visualisation process, a taxonomy of these parameters is necessary as well as description mechanisms to specify them in a concrete context.

Another distinction can be made between assistance systems which provide recommendations or those which criticise the applied visualisation. In general, recommending systems follow a problem-driven approach and perform an evaluation of a base of visualisation guidelines under consideration of context-specific parameters. Visualisation guidelines describe the knowledge or experiences of the effectiveness of visualisations for certain problems fields. Until now, the
visualisation guidelines applied in applications such as are APT ([Mack 86]), VISTA ([SeIg 90]), and IBIS ([SeFe 91]) were generally based on rules describing effectiveness aspects of visualisations which were drawn from experience and empirical tests.

A different approach is to provide support for a user in form of criticism of the applied visualisation pipeline and visualisation mappings, where the criticism can be informative or constructive. The general concepts are very similar to the ones discussed before if the criticism is generated based on visualisation guidelines. However, a much further going criticism can be produced based on a comparison of the data to be visualised and the information in the resulting image, either for the original data set or a test data set of known content. This corresponds to the evaluation of the presentation quality using measurement methods and applying an objective observer approach ([GeMü 94]). This aspect is discussed in some more detail in section 3.3 An example for a system following this approach is GrafikDesigner, which includes a beautifier to optimise 2D graphics generated in an editor ([Rome 94]).

The approach of a recommending assistant can easily be united with the criticising approach. Thus, the evaluation of the resulting image may change the recommendation or optimise the recommended visualisation mapping. Figure 1 presents a model of a problem-driven visualisation support system combining both approaches in such a way.

![Figure 1: Problem-driven visualisation support system](image-url)
In the following sections we will discuss the most important aspects of our visualisation assistant Vis-Wizz which follows the approach to give recommendation of a visualisation based on data characteristics and user’s goal as well as on a quantified evaluation of the visual representation to be generated. Moreover we will discuss some aspects with respect to an evaluation of the generated visualisation by a quality evaluation system.

3 A new approach to a problem-driven visualisation

What is meant by a problem-driven visualisation? Assuming that the underlying visualisation system offers several visualisation techniques, it has to be decided which technique to apply and how. To select an appropriate technique requires not only to know about size and type of data, but also to be aware of the goal the scientist has in mind when looking at the graphic presentation. Moreover, conventions of the underlying application contribute to the decision as well as characteristics of the hard- and software used.

Considering all these influences means to make the selection and realisation of a visualisation method problem-dependent. Consequently, a description of the underlying visualisation problem and a mechanism for selecting a suitable technique are needed. Based on this, the picked technique can be realised in an application builder by automatically generating the corresponding visualisation network. The automatic network generation is one way to integrate the support tool outlined below into a visualisation environment like shown in figure 2. We will look in more detail at the selection process and the influencing factors now.
3.1 Data specification and selection

Considering several approaches to data specification published in different contexts ([BeGr 89], [Robe 90], [Brod 92], [Brod 93]) and analysing a variety of visualisation applications, some data characteristics proved to be essential to the selection decision.

These are e.g. the type of data and the nature of its range (Is it continuous or discrete? Does an inherent order exist?). Furthermore it is of interest whether the data are defined over a spatial or temporal domain. If so, it is important to know about dimensionality and shape of grid and about the nature of domain (e.g. whether the data are defined point-by-point or over regions of the domain).

Another important factor is the information contained in the data. By the application of information theoretical methods, such as described by Theisel ([Thei 95]), one can decide which data contains relevant information and should be represented, which data is redundant and does not need to be shown since it does not contain additional information, and which data should be shown in combination to visualise their correlation.

The specification of these characteristics determines a basic set of techniques being potentially applicable to the given data set. This may lead, for instance, to a set of initial visualisation techniques for volume data, for vector fields, for scattered or for molecular data.

Hence the specified data characteristics permit a preliminary decision in the sense of a corresponding classification of visualisation techniques, but they can not provide a clear decision which technique to apply. Therefore we have to look at further criteria in order to reduce the set of data specific techniques to a set of techniques, which are well suited for the demands of the user, or even to the best suited one.

3.2 Specification of interpretation aims

When selecting a visual representation of a given data set one has to consider the intention of the data analyst. It makes a difference whether the focus is on frequencies and distributions, or on special values or on correlation between data variables, since the various techniques support these goals to a different degree.

There are two alternatives for the specification of interpretation aims. The one is to choose from an enumeration of verbally given aims, the other is to describe the aims in an abstract or formal
manner. The problem is to describe the aims in a way the selection mechanism is able to pro-
cess and the user is able to express. So one can define interpretation aims by means of first
order logic ([Thei 94]), but that does not meet the demand of the user for an application oriented
terminology. The user is expected to prefer descriptions like "diagnosis of a tumour’s size and
shape" rather than a first order logic expression.

Consequently, we have to map the application and data dependent expression of the user’s
intentions to a more abstract specification required by a selection mechanism.

3.3 Evaluation of visual representations

The evaluation of the quality of a visual presentation is not an easy problem. For a visual
presentation intended for human perception the final judgement depends on the observer and the
accomplishment of the communicative goal of the visualisation ([GeMü 94]). Thus, the quality
can only be evaluated with a good understanding of the observer’s visual capabilities, his
knowledge and his subjective feelings. Therefore, it can be argued whether an exact evaluation
of a visual representation may be achieved at all. However, even the application of limited
models for quality evaluation may still lead to a valuable support for the user.

As described before, two quality evaluation approaches can be distinguished in this context.
First, image quality can be evaluated by checking the applied visualisation method and
mappings against a set of general visualisation guidelines. This corresponds to an a priori
estimation of the expected visualisation quality. Thus, the effectiveness of the visual
presentation is determined according to certain characteristics of the direct mapping of infor-
mation in the visualisation pipeline. The advantage of such an approach lies in available
foundations in form of visualisation guidelines, such as discussed in the previous chapters.
However, a system of heuristic visualisation guidelines can hardly describe a closed and
consistent model of all aspects of the visualisation process, and, hence, can only handle a
limited amount of situations. Especially, intrinsic mappings of a rendering process and its
possible problems are seldom covered by visualisation rules and, thus, cannot be controlled.

Different from this approach, image quality can be evaluated a posteriori, and here two methods
can be distinguished: subjective tests, evaluated for instance with the ROC Method ([BeMD
93]), and objective measurement methods. Until now, subjective tests promise the best results
for quality evaluation of visualisation methods and parameter mappings. However, the results
of such tests are very domain and task specific and can seldom be generalised. Therefore, they
can only be applied to evaluate the effectiveness of a single technique in a specific application
context. Moreover, the test procedure is generally very time-consuming.
Consequently, the application of objective test methods seems to be the more appropriate approach on the way to an intelligent visualisation assistance system. Objective measurement methods try either to extract and measure physical entities from an image and to set them into relation with human quality evaluation, or they try directly to apply computational models of human vision to simulate an artificial observer. The information perceived by the artificial observer can be compared to the original information and the distortion can be measured. The results of this procedure may be transformed into a criticism of the visual presentation. A drawback of this approach is the complexity of the measurement methods and, until now, the missing of accepted models for human processing of visual information.

When applying a perception based approach to objective image quality measurement one has to understand that man analyses visual information on different layers of abstraction ([Marr 82]). Consequently, to model and to match human performance in quality evaluation it is necessary to use a similar approach and to distinguish between different aspects of the visualisation or different representation layers of visual information, such as an iconic, graphical and symbolic layer, and to measure the distortions on these layers ([GeMü 94]). Such a view of the problem corresponds to a Visual Computing approach ([HiMü 94]).

### 3.4 Automatic selection of visualisation methods

As described above the data specification leads to a first reduction of the set of known visual representations. But our research aims at finding those techniques, which meet the demands of the user discussed in the previous sections. Therefore we have to consider both the capabilities of each visualisation technique suitable for the given data set and the requirements arising from the underlying application problem.

To express a technique’s capabilities we apply a functional approach to the evaluation of the generated visual representation. For each technique and each interpretation aim it was analysed, which characteristics of the data influence the technique’s suitability for the given aim and in which manner. The influences are quantified ([Luko 93]) by considering the evaluation methods described above, visualisation guidelines published so far and threshold values established by interviews. This results in a set of functions, each of it determining a value which measures the suitability of a technique for a special aim with respect to the characteristics of the data set. For each technique a suitability vector is calculated containing a value for each interpretation aim.
On the other hand, the demands on the visual representation of the data set have to be specified. Therefore the user is asked to establish priority values for each interpretation aim, which leads to a priority vector.

Comparing the priority vector to the suitability vector of a technique results in a value measuring the support, which the user gets on his visualisation problem when utilising this technique. Starting from this measure and taking into account a certain determined threshold, the set of techniques is divided into two subsets containing those techniques, which do or do not meet the user’s demands, respectively. Consequently, if no technique exceeds the threshold, there is no appropriate technique available, which satisfies all the needs of the user. In the following section we will discuss this case in more detail.

4 The visualisation support tool

During our research on the selection process we developed the visualisation assistance tool Vis-Wizz in order to implement and test the selection mechanism discussed above. Starting from a short description of a typical session with our tool we will go into some interesting aspects in more detail.

The first step is to load the data set. As we deal with multi-dimensional data and plan to extend on further classes of scientific data, we chose netCDF ([Rew 90]) as the basic format for data import and export of our tool.

Vis-Wizz then extracts some characteristics from the data set, such as the type of values or the size of its domain, and analyses the data set based on information theory ([Thei 95]) to get meta data about the inherent information and the correlation between data variables. The user may interactively complete the data specification. For example, he can divide the range of a data variable into various intervals. Furthermore a data selection is performed by interactively choosing data variables and records to be visualised. Figure 3 presents the part of Vis-Wizz responsible for the selection of data variables. Interacting with the lower window the user may pick a data variable and ask for correlating variables in order to found his selection on it.

In the next step the user weights his demands by establishing priority values for each interpretation aim. At this stage the user may request for a suitable visualisation technique causing the assistance tool to generate a recommendation. Currently, Vis-Wizz handles a set of 15 techniques available for multi-dimensional data. For each of them it matches the suitability vector with the specified priority vector as described above. The threshold for the conformity, which decides whether a technique is suitable for the given problem, is interactively adjustable by the user.
If no technique can be found, which were able to satisfy the specified demands, Vis-Wizz offers two ways for a conflict solution. The first is to interactively re-specify the visualisation problem. For that, Vis-Wizz determines, which demands of the user cause the insufficient conformity of the application’s requirements and the technique’s capabilities. For each technique it displays the contradictions between its suitability for each aim and the priority given by the user. Furthermore, it estimates, whether the data set to be visualised should be reduced in order to improve the suitability of a technique and by which amount. Figure 4 shows that part of Vis-Wizz. Black bars represent the need to re-weight the corresponding interpretation aim, dark grey bars indicate the need to reduce the number of data variables or records by a certain amount. The user is free to weight the interpretation aims or to select a data subset again.

Figure 3: Selection of data variables
and to ask for a new recommendation. The disadvantage of this approach is that the user has to cut his original demands.

![Figure 4: Interactive conflict solution](image)

As this may be unacceptable to the user, Vis-Wizz provides an alternative way keeping the original weight of the visualisation problem. Since no technique meets the combination of all the user’s demands on its own, it is necessary to apply a variety of techniques, each of which satisfies a certain subset of the demands. Vis-Wizz seeks for a minimum set of techniques in order to keep down the number of visual representations to be examined in the end. However, it is left to the user to lead all the partial information, carried by the set of images, into general information about his problem. Figure 5 shows how Vis-Wizz presents its recommendation.

Regardless of the way the session went so far, now at least one visualisation technique is offered to the user. He may ask the assistance tool to visualise the specified data by the proposed techniques. Vis-Wizz causes the general purpose visualisation system IRIS Explorer to start and to create the appropriate visualisation network. This is done by automatically generating scripts from script fragments, which describe special visualisation networks in a generic manner, using meta information about the visualisation problem and the proposed techniques as parameters. After having processed the script, the whole functionality of IRIS
Explorer is available to the user. The integration with other visualisation systems is possible, if only a scripting language is provided, which allows to control the network generating process.

Figure 5: Recommendation of a combination of techniques

During the whole session the user may ask for help on several topics, like on the problem specification, on the selection process or on a special visualisation technique. Vis-Wizz offers an on-line help with explanatory texts and starts IRIS Explorer with an example visualisation network in order to provide an impression of a special visual representation if desired.

5 Conclusions and future work

We developed an visualisation assistance tool, which represents under various aspects a new approach to user support during the visualisation process. First of all, Vis-Wizz considers the quantitative evaluation of visualisation guidelines to be the fundamental concept for the recommendation of visualisation methods and mappings. This is realised by a functional approach to the evaluation of visual representations using data characteristics as parameters.
Moreover, Vis-Wizz works problem-driven as it starts from a description of the visualisation problem including data characteristics as well as the visualisation goals.

Another important aspect of Vis-Wizz is its integration into a standard visualisation system. This allows to use the concepts and the functionality of a well-tested system and prevents the user from the difficulties of getting into yet another visualisation tool. The next step of our research is to join together our assistance tool and a database management system as a further component of a visualisation environment.

A limitation of Vis-Wizz is its specialisation to the visualisation of multi-dimensional data. Since a lot of data has a strong relation to time and space the integration of further data-classes into Vis-Wizz is necessary. Consequently, current work is concerned with the extension of Vis-Wizz by data classes for volume data and time dependent data.

Acknowledgement

This work is supported by the Deutsche Forschungsgemeinschaft (DFG).

References


