标题：可视化的价值
文件：2005_Value_Visualization_annotated.pdf

注解
这些项目符号上的数字与添加到研究论文上的数字相对应。

1. 可视化领域正在成熟。解决了许多问题，并寻求新的方向。为了做出正确的选择，需要了解虚拟化的目的和意义。尤其是，如果我们可以评估什么是好的可视化效果，那就太好了。
2. 现代社会面临着数据爆炸。 MRI扫描仪等采集设备，超级计算机上的大规模仿真以及证券交易所的股票交易都产生大量数据。数据的可视化使研究人员，分析师，工程师和非专业观众有可能以有效的方式获得对这些数据的见解，这要归功于人类视觉系统的独特功能，使我们能够检测有趣的特征和模式在短时间内。
3. 作为研究人员，必须做出决定，从花时间在哪个领域上到实施哪种特定解决方案有关；作为指导者，必须向学生提供指导；作为审稿人，必须对新的结果和新研究的建议进行判断，如果值得发表或提供资金，则应征求他们的意见；作为一家初创公司的顾问，必须发现新颖且有利可图的方向。
4. 可视化必须有效。换句话说，可视化应该执行其应做的工作，并且必须使用最少的资源来执行此操作。
5. (Lei，“背景”部分提供了有用的历史信息)
6. 图1显示了基本模型。方框表示容器，圆圈表示将输入转换为输出的过程。
7. 洞察力是可视化的传统目标。这个词本身很棒，暗示了对科学进步的高水平贡献。使用户能够看到他们不知道的事物，并且这种见解有助于他们定义新的问题，假设和数据模型。
8. 用户有问题，他必须决定要采取的动作，并要做出该决定，他需要信息。可视化应该使他能够从数据中提取相关信息。
9. 决策通常是关于是否要采取的行动。例如，应该买卖股票，是否要经营病人，组织中的哪些人适合晋升等。因此，我建议我的学生在使用后搜索并列举用户可能采取的行动他们的预期工具。如果无法找到或定义此类动作，则可视化的价值值得怀疑。如果我们想提供附加价值，仅声称可视化能够提供洞察力还不够。
10. 当存在自动方法来提取相关信息时，可视化是没有用的。
11. 集成自动化方法（例如来自统计或数据挖掘的方法）以及可视化是一个好主意
12. 让我们考虑一个潜在的可视化客户，例如研究人员进行复杂的模拟。首先，他必须意识到也许可视化可以帮助他理解他的数据。
13. 研究人员必须熟悉可能的解决方案。不幸的是，没有介绍和比较新颖可视化技术的书籍。
14. 如果没有更好的替代方法来检查复杂数据，则必须使用可视化。另一道防线是，可视化不应用于验证最终的真实性，而应启发新的假设，随后再进行检查。
15. 因此，新方法的开发人员应仔细考虑良好的默认设置或自动设置可视化参数的方

法，以便尽可能多地传播知识。

16. 交互可以极大地增强您对数据的理解。最重要的情况是显示的数据量不适合屏幕显示，或者太大而无法从单个图像中理解。在这种情况下，必须支持数据的导航和选择
17. 几个按钮下应有适当的预设，以方便可视化。
18. 可视化的主要用例是探索（用户不知道数据中的内容）和表示（某些结果必须传达给其他人）。
19. 从广义上讲，我们可以在任何地方查看可视化。商业电视使用可视化显示新化妆品的化学奇迹，吸尘器的独创性以及为什么新的健身设备不会伤害您的背部。
20. 我们可以将自己视为一组难题解决者，而面临的挑战是开发新的，简单的，优雅的解决方案，从而为我们所有人提供智力和美学上的满足。
21. 除了将可视化视为一项技术或一门艺术之外，我们还可以将可视化研究视为一门科学学科。
22. (与该主题相关的研究论文)

The Value of Visualization

Jarke J. van Wijk*
Dept. Mathematics and Computer Science
Technische Universiteit Eindhoven

ABSTRACT

1 The field of Visualization is getting mature. Many problems have been solved, and new directions are sought for. In order to make good choices, an understanding of the purpose and meaning of visualization is needed. Especially, it would be nice if we could assess what a good visualization is. In this paper an attempt is made to determine the value of visualization. A technological viewpoint is adopted, where the value of visualization is measured based on effectiveness and efficiency. An economic model of visualization is presented, and benefits and costs are established. Next, consequences for and limitations of visualization are discussed (including the use of alternative methods, high initial costs, subjectiveness, and the role of interaction), as well as examples of the use of the model for the judgement of existing classes of methods and understanding why they are or are not used in practice. Furthermore, two alternative views on visualization are presented and discussed: viewing visualization as an art or as a scientific discipline. Implications and future directions are identified.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces; I.3.6 [Computer Graphics]: Methodology and Techniques I.3.8 [Computer Graphics]: Applications

Keywords: Visualization, evaluation

1 INTRODUCTION

2 Modern society is confronted with a data explosion. Acquisition devices like MRI-scanners, large scale simulations on supercomputers, but also stock trading at stock exchanges produce very large amounts of data. Visualization of data makes it possible for researchers, analysts, engineers, and the lay audience to obtain insight in these data in an efficient and effective way, thanks to the unique capabilities of the human visual system, which enables us to detect interesting features and patterns in short time.

Many of us will have written paragraphs like the preceding one, where I attempted to give the standard rationale of our field. In 1987, when the influential ViSC report [16] of the NSF appeared, the expectations were high. Visualization was considered as vital and highly promising for the scientific process. Nowadays, much progress has been made. The advances in graphics hardware are astonishing, most laptop computers are graphics superworkstations according to the standards of just a decade ago. Many new methods, techniques, and systems have been developed. Some of them, such as slices, height-surfaces, and iso-surfaces are now routinely used in practice.

On the other hand, many of these new methods are not used in real-world situations, many research results are nowadays considered as incremental by reviewers, and our prospective users rarely go to our conferences. So, are we, as researchers in visualization, on the right track?

In this paper I want to give a contribution to the discussion on the status and possible directions of our field. Rather than to pinpoint specific topics and activities, my aim is to detect overall patterns, and to find a way to understand and qualify visualization in general. This is an ambitious and vague plan, although the basic ground for this is highly practical.

I have to make decisions on visualization in many roles. As a researcher, decisions have to be made ranging from which area to spend time on to which particular solution to implement; as a supervisor, guidance to students must be provided; as a reviewer, new results and proposals for new research must be judged, and opinions 3 are expected if they are worth publishing or funding; as advisor in a start-up company, novel and profitable directions must be spotted. All these cases imply judgement of the value of visualization in varying senses.

How to assess the value of visualization? Visualization itself is an ambiguous term. It can refer to the research discipline, to a technology, to a specific technique, or to the visual result. If visualization is considered as a technology, i.e., as a collection of methods, techniques, and tools developed and applied to satisfy a need, then standard measures apply: Visualization has to be *effective* and *efficient*. In other words, visualization should do what it is supposed 4 to do, and has to do this using a minimal amount of resources. One immediate and obvious implication is that we cannot judge visualization on its own, but have to take into account the context in which it is used .

In section 2 a short overview is given of the background of the topic discussed here. In section 3 an economic model of visualization is proposed. The basic elements are identified first, the associated costs and gains are added next. Various implications of the model are discussed in section 4. In section 5 this model is applied to several cases. In section 6 the model is discussed and alternative views are considered, followed by conclusions in section 7.

Finally, this topic is on one hand very general, high-level, and abstract; on the other hand, it is also very personal, in the sense that it is about values (which are subjective), and valuation of ones own work. To reflect this, I use the first person in this paper, to emphasize that the opinions given are personal. Most examples I use come from my own work, often done together with coworkers. The main reason for this is simply that I am most familiar with it, not only with the techniques and results, but also with the context in which it took place.

2 BACKGROUND 5

If we use 1987 as the year where visualization started, our discipline celebrates this year its 18th anniversary. In the Netherlands, at this age a person is considered mature. Many things have changed since 1987. Graphics hardware developments are amazing, as well as the large amount of techniques that have been developed to visualize data in a variety of ways.

There are signals that there is a need to reconsider visualization. First of all, there seems to be a growing gap between the research community and its prospective users. Few, if no attendants at the IEEE Visualization conference are prospective users looking for new ways to visualize their data and solve their problems. Secondly, the community itself is getting both more specialized and

*e-mail: vanwijk@win.tue.nl

IEEE Visualization 2005
October 23-28, Minneapolis, MN, USA
0-7803-9462-3/05/$20.00 ©2005 IEEE.

critical, judging from my experience as paper co-chair for IEEE Visualization 2003 and 2004. In the early nineties, the field lay fallow, and it was relatively easy to come up with new ideas. The proceedings in the early nineties show a great diversity. Nowadays the field is getting more specialized, submitted work consists often of incremental results. This could signal that our field is getting mature. On the other hand, it is not always clear that these incremental contributions have merit, and reviewers are getting more and more critical. Thirdly, some big problems have been solved more or less [14]. For volume rendering of medical data sophisticated industrial packages that satisfy the needs of many users are available.

These trends urge a need to reconsider the field, and to think about new directions. Several researchers have presented [7, 9, 17] overviews of current challenges. Another great overview of the current status of visualization and suggestions for new directions is provided by the position papers [3] contributed by the attendants of the joint NSF-NIH Fall 2004 Workshop on Visualization Research Challenges, organized by Terry Yoo. Many issues are mentioned several times, including handling of complex and large data sets, uncertainty, validation, integration with the processes of the user, and a better understanding of the visualization process itself. One particularly impressive and disturbing contribution is [14], for its title, the name and fame of the author, and the vivid description that indeed the field has changed and new directions are needed.

In this paper no attempt is made to summarize or overview these challenges, but the aim is to find a model or procedure to judge in general if a method is worthwhile or not. In the following sections, a first step towards such a model is presented. Much of it is evident and obvious. As a defense, some open doors cannot be kicked open often enough, and also, if obvious results would not come out, the model and the underlying reasoning would be doubtful. Some statements made are more surprising and sometimes contrary to main stream thinking. To stimulate the debate, I have taken the liberty to present these more extreme positions also, hoping that some readers will not be offended too much.

3 Model

In this section a generic model on visualization is proposed. First, the major ingredients are identified; secondly, costs and gains are associated. The model is abstract and coarse, but it can be used to identify some aspects, patterns and trends.

3.1 Visualization and its context

Figure 1 shows the basic model. Boxes denote containers, circles denote processes that transform inputs into outputs. The aim here is not to position different visualization methods, for which a taxonomy would be a more suitable approach, but rather to describe the context in which visualization operates. No distinction is made, for instance, between scientific visualization and information visualization, at this level there is much more they share than what separates them.

In the following we describe the various steps. We use a mathematical notation for this, merely as a concise shorthand and to give a sense of quantification than as an exact and precise description. Processes are defined as functions, but the domains and ranges of these are ill-defined.

The central process in the model is visualization V:

$$I(t) = V(D, S, t).$$

Data D is transformed according to a specification S into a time varying image $I(t)$. All these should be considered in the broadest sense. The type of data D to be visualized can vary from a single bit to a time-varying 3D tensor field; the specification S includes a specification of the hardware used, the algorithms to be applied

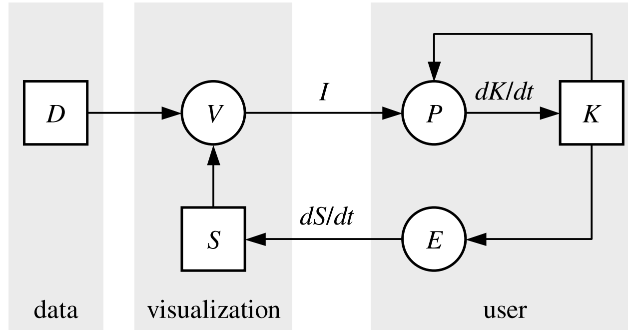

Figure 1: A simple model of visualization

(in the form of a selection of a predefined method or in the form of code), and the specific parameters to be used; the image I will often be an image in the usual sense, but it can also be an animation, or auditory or haptic feedback. In other words, this broad definition encompasses both a humble LED on an electronic device that visualizes whether the device is on or off, as well as a large virtual reality set-up to visualize the physical and chemical processes in the atmosphere. The image I is perceived by a user, with an increase in knowledge K as a result:

$$\frac{dK}{dt} = P(I, K).$$

The amount of knowledge gained depends on the image, the current knowledge of the user, and the particular properties of the perception and cognition P of the user. Concerning the influence of K, a physician will be able to extract more information from a medical image than a lay-person. But also, when already much knowledge is available, the additional knowledge shown in an image can be low. A map showing the provinces of the Netherlands provides more new information to a person from the US than to a Dutch person. Also, the additional value of an image of time-step 321 is probably small when time-step 320 has been studied just before. Concerning the influence of P, a simple but important example is that a color-blind person will be less effective in extracting knowledge from a colorful image than a person with full vision. But also, some people are much better than others in spotting special patterns, structures, and configurations.

The current knowledge $K(t)$ follows from integration over time

$$K(t) = K_0 + \int_0^t P(I, K, t)\, dt$$

where K_0 is the initial knowledge.

An important aspect is interactive exploration, here represented by $E(K)$. The user may decide to adapt the specification of the visualization, based on his current knowledge, in order to explore the data further

$$\frac{dS}{dt} = E(K),$$

hence the current specification $S(t)$ follows from integration over time

$$S(t) = S_0 + \int_0^t E(K)\, dt$$

where S_0 is the initial specification.

3.2 Economic Model

To assess if a visualization method is worthwhile, we must assess its value. We propose to use profitability in an economic sense as a measure for this. We simplify this by assuming that there is a homogeneous user community, consisting of n users which use a certain visualization V to visualize a data set m times each, where each session takes k explorative steps and time T. This is a crude simplification of course. In the real world, the user community will often be highly varied, with different K_0's and also with different aims. The costs associated with using V come at four different levels:

- $C_i(S_0)$: *Initial development costs*. The visualization method has to be developed and implemented, possibly new hardware has to be acquired.
- $C_u(S_0)$: *Initial costs per user*. The user has to spend time on selection and acquisition of V, understanding how to use it, and tailoring it to his particular needs.
- $C_s(S_0)$: *Initial costs per session*. Data have to be converted, and an initial specification of the visualization has to be made.
- C_e: *Perception and exploration costs*. The user has to spend time to watch the visualization and understand it, as well as in modification and tuning of the specification, thereby exploring the data set.

The total costs are now given by

$$C = C_i + nC_u + nmC_s + nmkC_e.$$

The return on these investments consists of the value $W(\Delta K)$ of the acquired knowledge $\Delta K = K(T) - K(0)$ per session, multiplied by the total number of sessions:

$$G = nmW(\Delta K)$$

and hence for the total profit $F = G - C$ we find

$$F = nm(W(\Delta K) - C_s - kC_e) - C_i - nC_u.$$

This gives us a recipe to decide on the value of a visualization method. Positive are high values for n, m, $W(\Delta K)$, and low values for C_s, C_e, C_i, C_u, and k. Or, in other words, a great visualization method is used by many people, who use it routinely to obtain highly valuable knowledge, without having to spend time and money on hardware, software, and effort. Indeed, quite obvious.

4 Implications

Quantification of the elements of the model is hard. In this section we discuss this in more detail, as well as a number of other issues implied by this model.

4.1 Valuable knowledge

Insight is the traditional aim of visualization. The term itself is great, and suggests a high-level contribution to the advance of sci-
7 ence. Users are enabled to see things they were not aware of, and this insight helps them to define new questions, hypotheses, and models of their data. However, from an operational point of view, the term insight does not help us much further to assess the value of visualization. One problem is that we cannot directly observe or measure how much insight is acquired, and also, it is difficult to assess what the value of that insight is. In the model we use the term knowledge, but this suffers from the same limitations. Also, there is a strange paradox in the basic paradigm of visualization. We don't know what information is contained in the data, hence we make pictures to get insight. But if we do not know which specific aspects or features should be visible, we cannot assess if we are successful or not.

Nevertheless, we should try to measure or estimate $W(\Delta K)$, if we want to assess the value of visualization, especially because it is the only term in the model for F with a positive sign. An operational approach is to consider the use of visualization as an element in problem solving. The user has a problem, he must decide which
action to take, and to make that decision he needs information. The 8
visualization should enable him to extract the relevant information from the data.

Decisions are typically about actions to be taken or not. For instance, should a stock be bought or sold, should a patient be operated or not, which people in an organization are candidates for promotion, etc. Hence, I recommend my students to search for and enumerate possible actions of users after using their prospec-
tive tools. If such actions cannot be found or defined, the value of 9
visualization is doubtful. Just claiming that a visualization gives insight is not enough, if we want to offer additional value.

If we know to which actions the visualization should lead to, the next steps are assessment whether the knowledge derived from the visualization does indeed support the decision, and also, to assess the economic value of this decision. This is not easy, but one can try for instance to estimate how much time is saved, or try to quantify the consequences of a wrong decision.

4.2 Alternative methods

Efficiency is relative, an aspect that is not captured explicitly in the model. One could predict a high value for F for a new method, however, if other methods are available to obtain the same knowledge against lower costs, then very likely the value for n is overestimated. Or, stated simply, if a better solution already exists, nobody will use the newer one. The model is too simple here. The effective value of n itself is not a parameter, but a function of, among others, the perceived benefit by potential users.

Developers of new visualization methods should be aware of alternative solutions, and carefully study their advantages and limitations. New methods are not better by definition. Especially when existing methods are heavily used in practice, they have proven to have value. It is often hard to beat straightforward solutions; for instance, in many cases just using a line graph is the best way to show a time-varying signal.

A defense often heard for a lesser performance of new methods compared to existing ones is that the users have not had enough time to get accustomed to them. In some cases this might hold, but an equally viable hypothesis is that an existing method is simply better. For instance, just showing a set of objects in a list enables linear scanning, whereas scanning a fancy 2D or 3D display where the objects are distributed over space is much harder [18].

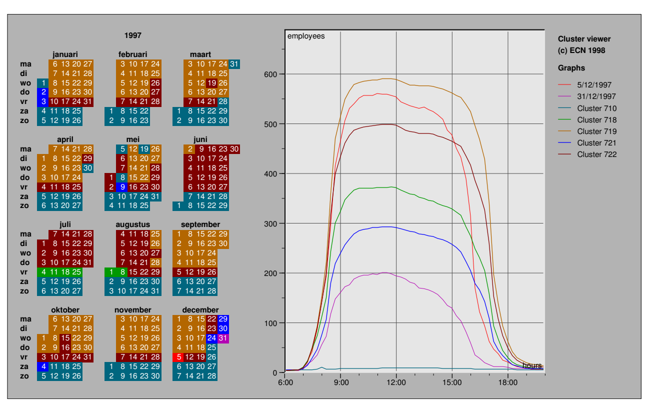

Figure 2: Visualization of daily patterns [28], an example of the combined use of conventional statistical and graphical methods.

Alternative methods are not limited to visualization methods. For instance, when an automatic method exists to extract the rel-
10 evant information, visualization is useless. Visualization is not 'good' by definition, developers of new methods have to make clear why the information sought cannot be extracted automatically. One reason could be that such automated methods are not fullproof. In this case, integration of automated methods, for instance from
11 statistics or data-mining, and visualization is a great idea, see for instance the work underway and led by Jim Thomas in the Visual Analytics arena [19].

Figure 2 shows an example where we used standard methods in a new combination [28]. For the analysis of a time-series of one year, daily patterns were clustered, i.e., finding similar daily patterns was automated. The results are shown using two conventional representations: average daily patterns of clusters are shown as graphs, and the days per cluster are shown on a calendar. The approach is straightforward and conventional, and very effective.

4.3 High initial costs

One important reason that new visualization techniques are not used in practice is the high initial cost per user $C_u(S_0)$ involved. Let us
12 consider a potential customer for visualization, for instance a researcher doing complex simulations. First, he has to realize that maybe visualization can help him to understand his data. This is not obvious, he already uses some methods to extract information from his results in a condensed form. For instance in molecular dynamic simulations, one typical aim is to derive large scale quantities (temperatures, porosity, etc.) via simulation from the properties on a small scale (size of ions, fields, etc.). Such large scale quantities can be calculated fairly easily from the raw data. Mathematicians working in Computational Fluid Dynamics are often not interested in particular flow patterns, but rather in convergence of numerical methods and conservation of quantities, which again can be calculated easily and summarized in a few numbers.

The easiest way to visualize data is to use post-processing capabilities that are integrated with the software used. Commercial packages for, for instance, computational fluid dynamics or finite element simulation offer these. From a visualization point of view, the techniques offered are far from state of the art: Usually just options like iso-surfaces, color mapping, slicing, streamlines and arrow plots are provided. But if these meet the demands of our user, then this is a highly cost-effective way.

Suppose that this option is not available or falls short. The next step is to find alternatives. Our researcher has to get acquainted with
13 possible solutions. Unfortunately, there are no books that present and compare novel visualization techniques (like volume rendering or topology based flow visualization) at an introductory level. So he has to study research papers, or search and get in contact with an expert in the field.

Next steps are also costly. Maybe he can get a research prototype to work with, or else he has to (or let somebody) implement the novel techniques. Often additional software has to be developed to convert his data to a suitable format.

This all takes much time and effort, while it is unclear whether the new method will indeed solve his problem. Hence, a rational decision is to abstain from this.

There are of course ways to share the initial costs with others. A group of researchers can take advantage of an initial investment by one of them. Also, providers of simulation software can be asked to integrate new methods. Visualization does not seem to have a high priority here however. For an impression of what providers think to be important for their customers, we can have a look at web-sites of companies like MSC or Fluent, and observe that features like advanced simulation capabilities and tight integration are promoted much more than visualization, which is just mentioned in passing by under the header of post-processing.

4.4 Visualization is subjective

In the ideal case, one would hope that extraction of knowledge from data is an objective process, in the sense that the outcome does not depend on who performs it, and that the analysis can be repeated afterwards by others, with the same outcome. Statistics aims at this, a typical pattern is the use of statistical tests to validate hypotheses on the data. Such tests make assumptions on the data (such as a normal distribution) and have free parameters (like the confidence level), but furthermore, they do meet the criteria for objectiveness.

Unfortunately, visualization often does not meet this aim. Consider

$$\frac{dK}{dt} = P(V(D,S,t),K).$$

This simply means that the increase in knowledge using visualization not only depends on the data itself, but also on the specification (for instance, which hardware has been used, which algorithm has been used and which parameters), the perceptual skills of the observer, and the a priori knowledge of the observer. Hence, the statement that visualization shows that a certain phenomenon occurs is doubtful and subjective.

An even harder case is the statement that a certain phenomenon does not occur. I have often spent hours visualizing data, searching for patterns and structure. Sometimes some result could be produced using a particular setting of the parameters, in other cases I failed to do so. When a visualization does not show clear patterns, it is hard to decide if this is a limitation of the visualization method, or that the setting of the parameters was wrong, or that the data simply does not contain significant patterns.

This does not mean that visualization is useless. If there are
no better alternatives to inspect complex data, visualization has to 14
be used. Another line of defense is that visualization should not be used to verify the final truth, but rather to inspire to new hypotheses, to be checked afterwards. Part of the subjectiveness can be eliminated by simply showing the visualization to the audience, so that they can view and judge it themselves. However, this does not take away the subjectiveness inherent in S, as a second hand viewer we do not know how sensitive the ultimate visualization is to changes in scales and/or selections of the data.

4.5 Negative knowledge

In the previous subsection we considered subjective aspects of visualization. There is another problem: Visualizations can be wrong and misleading. Or, in the terminology introduced here, negative knowledge ($|\Delta K| < 0$) can be produced. Tufte has introduced the *lie-factor* [23], which he defined as the ratio of the size of an effect shown in the graphic to the size of the effect in the data.

Here, I just want to give an example of my own experience with this. A long time ago I visualized the waves produced by ships for a maritime research institute. The data were the result of simulations. Figure 3 (a) shows the result of bilinear interpolation of the data. I found these results unclear, hence I decided to use an interpolating spline, thereby smoothing the surface while remaining faithful to the data. Figure 3 (b) shows clearly that two sets of waves are generated: the standard waves as well as a set of waves orthogonal to this. I proudly presented this discovery to the researcher, who immediately replied that this was physically totally impossible. A much better visualization is shown in figure 3 (c), where an approximating spline is used. The artifacts in the middle image are the result of aliasing. The data orthogonal to the ship are sampled close to the Nyquist frequency, interpolation gives rise to aliases, which corresponding waves have in this 2D case a different direction than the original wave. A smoothing interpolating spline smoothes away the high frequencies, but the first aliases survive and give rise to wrong interpretations. I learned from this that interpolation is not by definition better than approximation, and also that the judgement

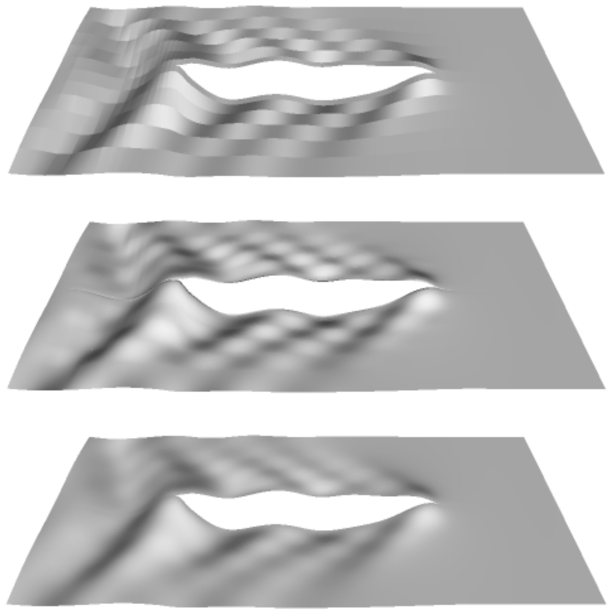

Figure 3: Wave surface, from top to bottom (a) bilinear interpolation, (b) cubic interpolation, (c) cubic approximation. Incorrect interpolation leads to artifacts.

of an expert, with a high K_0, is vital for proper interpretation and validation. I never published this, and also, articles on limitations and pitfalls of visualization are scarce. For an advancement of the field, more such reports would be highly beneficial.

4.6 Interaction

Interaction is generally considered as 'good'. One could advocate the opposite: Interaction should be avoided, and well for two reasons. First of all, as mentioned before, allowing the user to modify S freely will lead to subjectiveness. It is tempting to tune the mapping so that the desired result comes out strongly, but this can be misleading. Also, high customization can make it hard to compare different visualizations. Secondly, interaction is costly, and leads to a high C_e. Rerendering the image after a change of the mapping or the point of view taken requires often a few seconds, viewing it again also. If many options are available to modify the visualization, trying them all out can take hours. A developer of a new
15 method therefore should think carefully about good defaults, or automatic ways to set the visualization parameters, so that as much knowledge is transferred as possible.

16 Obviously, in many cases interaction strongly enhances the understanding of the data. The most important case is simply when the amount of data to be shown does not fit on the screen, or is too large to be understood from a single image. In this case, navigation and selection of the data has to be supported. Ideally, the user has to be provided with cues that will lead him quickly to images where something interesting can be seen. Another case is during development of new methods. I stimulate my students to make every aspect of their new methods customizable via user interface widgets, so that the total solution space can be explored. However, for the final versions of their prototypes I recommend them to offer
17 suitable presets under a few buttons, so that a good visualization can be obtained with little effort.

5 Examples

In this section a number of (classes of) techniques are considered and the cost model is used to explain their adoption in practice.

5.1 Texture based flow visualization

The use of texture to visualize fluid flow has been introduced in the early nineties. The idea is that dense textures enable viewers to judge the direction of flow at all locations of the plane, whereas the standard arrows and streamlines only give discrete and hard to interpret samples. The topic has been studied heavily in the visualization community, a recent non-exhaustive overview [13] has 90 references. The progress made in this decade is great. The early Spot Noise technique [24] was an interesting first attempt, in 1993 Cabral and Leedom introduced Line Integral Convolution (LIC), which gave high quality renderings of 2D fluid flow [5]. Many other variations and additions have been presented since then, for instance to handle flow on surfaces and in volumes, and also to boost the performance, using software or hardware acceleration [13]. Nowadays, high quality 2D texture images of flow fields can easily be generated on standard hardware at 50 or more frames per second [25]. This seems a success story, but on the other hand, these methods are not integrated in commercial software, users of Computational Fluid Dynamics (CFD) are typically completely unaware of their existence, let alone that they routinely use them to solve their problems. Here I use texture based flow visualization because I am most familiar with it, but for other classes of methods, such as topology based flow visualization and feature based flow visualization, similar patterns seem to apply.

How can we explain this? We consider the parameters of the cost model. The number of users n is not too great. CFD is vital for some areas, but there are few cases where CFD is routinely used for screening, compared to for instance medical applications. The frequency of use m is also not very high. Often, CFD-users spend much time on defining the model, simulations can also take a long time. By then, they are very familiar with their models (high K_0). For the analysis of the results many alternative options are available, including composite quantities (such as lift of an airfoil) and straightforward cross-sections and arrow plots, with low costs. The use of texture based visualization incurs at least a high value for C_u (see section 4.3). The additional ΔK that texture based visualization offers is unclear. Laidlaw et al. [12] have compared different vector visualization methods. LIC turned out to yield better results for critical point detection, but worse results for other aspects, such as estimation of the angle of the flow. Also, standard LIC does not give the sign of the direction of the flow. Hence, we can doubt about the value of ΔK. And finally, it is not clear what the real value is of this ΔK, in the sense that better visualization leads to better decisions. At least, so far there does not seem to be such a strong need for better visualization methods in the CFD community that they have attempted to integrate these methods into their packages.

5.2 Cushion treemaps

Also in the early nineties, Johnson and Shneiderman introduced the concept of a treemap [8] to visualize large hierarchical data sets. The base algorithm is straightforward: A rectangle is recursively subdivided according to the hierarchical data, in such a way that the size of each rectangle corresponds to the size of each leaf element. In the late nineties we proposed to use hierarchical cushions to show the underlying hierarchical structure more clearly [26]. We packaged this technique in 2000 in SequoiaView [1], a tool for the visualization of the contents of a hard disk (figure 4), and made this publicly available as freeware. Since then, SequoiaView has been downloaded about 400,000 times from our site. Also, it has been

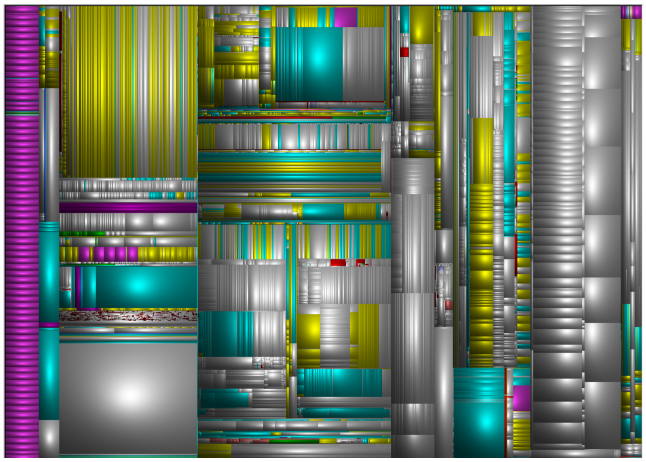

Figure 4: Visualization hard disk using SequoiaView [1, 26, 27], an example of an application that has found an audience.

distributed three times via CD with the German computer magazine C't. This is an example how visualization has reached an audience.

The economic model helps to explain this result. First, the number of (potential) users is very large, in principle equal to the number of PC users. Typically, such a tool is used several times per year, which is not very high, but not neglectable. Alternative solutions for this problem are scarce (SpaceMonger, using also treemaps is an example), and getting an overview of a hard disk is hard using Windows Explorer.

Information can be derived fairly easy from the visualization. It is easy to spot large files, large directories, and large collections of files. Furthermore, this information is directly valuable for the user: The tool can help (and many users have confirmed this) to delay buying a new hard disk. The action is clear here: removal of files. We offer an option to start up Windows Explorer from SequoiaView to remove files manually. The initial costs per user are low: The tool itself is freeware, it only has to be downloaded and installed. The costs per use case are minimal as well. By default, the tool starts to collect data from the last folder specified, and an image is shown automatically. Exploration is easy: Extra information per file can be obtained by hovering the pointer over the rectangles.

In summary, F is high in this case. We would like to think that this is a result of our visualization method, however, the main reasons are probably that our tool meets a real need of real users, and that the costs, in all respects, are minimal.

5.3 Presentation vs. exploration

18 Next we consider a more general case. The main use cases for visualization are exploration (where users do not know what is in the data), and presentation (where some result has to be communicated to others). It is hard to quantify this, but my impression is that many researchers in visualization consider exploration as the major raison d'être for visualization, whereas presentation is considered as something additional and not too serious. However, from my own experience, presentation is at least just as important as exploration. Many users find videos and images attractive for presenting their work at conferences; the popularity of visualization tools and demos often rises sharply just before open days. For years I had a pleasant and fruitful cooperation with Flomerics Ltd. in the UK. This company develops CFD-based tools for, amongst others, thermal assessment for the electronics industry. My major contact there was the marketing manager, who could use visualization to show the benefits of the CFD tools to managers.

In a broader sense, we can view visualization everywhere. Com- 19
mercial television uses visualization to show the chemical miracles of new cosmetics, the ingenuity of vacuum-cleaners, and why a new fitness device does not harm your back. Obviously, such visualizations are probably not the result of visualizing data, but rather the result of fantasy of advertisement agencies. Selling stuff is not only the realm of business, but also of science itself. Once I heard someone state: *The purpose of visualization is funding, not insight*. We can explain the value of visualization for presentation with the cost model. If we consider the viewers of such visualizations as the users, we see that n is high; K_0 is low (the viewers know little about the topic, so much can be gained); the action to be taken is clear (buy a product, fund research) and has direct economic consequences; the costs for the viewers are low (they just have to watch the visualization), although they can be high for the presenter. And furthermore, for these purposes there are almost no alternative or competing techniques. Pure facts (product X saves Y percent of time) can be convincing, but to make plausible why, and also to show that this is all Scientifically Sound, visualization is the way to go.

6 Discussion

In the preceding sections a number of questions were raised and various disturbing statements were made. There are many objections that can be made, and in this section some of them are given. One important distinction is to consider visualization either as technology, art, or as science. Associated with these are a number of routes for future work.

6.1 Technology

In the cost model, visualization is considered as a technology, to be measured for utility. In this context, research in visualization should lead to new solutions that are useful in practice. Not all the work done is successful in this respect, but we can find a number of reasons to explain this.

First of all, innovation is a merciless process, where only few new solutions survive. A rule of thumb in product development is that thousand ideas lead to hundred prototypes, which lead to ten products, out of which just one is successful. The visualization research community operates in the start of this pipeline, hence it should come as no surprise that not everything finds its way. We can see it as a mission to develop inspiring new ideas, which are a primary fuel in the innovation process.

Creativity however consists of two parts: creation of new ideas as well as selection of the best ones. The first task is fulfilled properly by the visualization community, the second is not. The number of careful validations of visualization methods is still low, although this seems to be improving in the last years.

Secondly, innovation is a long chain. Developing new methods is quite different from turning these into products and marketing them. There is a gap between our prospective users and the research community. Both do not have the proper stimuli to bridge this gap: individual researchers are too busy increasing the number of publications they are judged on, and for the end-users implementing new methods is far too costly. The gap can be filled in different ways. One way is via commercial companies (spin-off companies, or companies that integrate visualization in their simulation packages), an alternative is via open source and academic development and maintenance, funded by government agencies. VMD [2] is an example of the latter category. As a corollary, if we think that visualization is useful and that this gap causes the lack of adoption, we should aim at increasing funding for more practical activities. Or we should start up companies.

Thirdly, one could state that all this is a matter of time. It takes time before new ideas penetrate, before new users become aware of new methods, before initiatives are taken to integrate new methods into existing systems. This might be true in some cases, however, it is also too easy to use this as an excuse. It could be used for any method, hence it does not help us to distinguish between good and bad ones.

Fourthly, the focus in the model is on large numbers of users and use cases. One can also consider cases where the number of users is small, but where the value of the result is very large. In the books of Tufte some great cases are presented, such as Snow's discovery of the cause of a cholera epidemic in 1854 [21]. Are there recent cases for new visualization methods? Cases that enabled the researcher to obtain a major scientific insight, to save many lives, or to solve a crucial technological problem? One would like to read more case studies in this spirit, which show that visualization is worthwhile and can make a difference.

Finally, one defense is that maybe we are not doing too bad, compared to other disciplines. Many disciplines (for instance, in mathematics) do not care about practical usability at all, for some computer science fields that do claim to have practical relevance it is also hard to see the adoption in practice. Why should we bother? This notion is explored further in the next subsection.

6.2 Art

One could claim that visualization has value in its own right, and for its own purposes. One part of this is in the results: Some of the images we produce have a clear aesthetic value. But the art of visualization can also be found in the ideas, methods, and techniques developed. We can consider ourselves as a group of puzzle solvers,
20 and the challenge is to develop new, simple, and elegant solutions, which provide us all with intellectual and aesthetic satisfaction.

This is not a line of defense that can help us to convince our prospective users and sponsors. Nevertheless, I do want to mention it, because it can give a powerful thrust (and obviously also because results of this possibly will find applications in the real world). In the early nineties, I worked hard on using texture for visualization – not to satisfy users, but simply because the puzzle was tough, challenging, and hard to crack. The work of our student Ernst Kleiberg on botanically inspired tree visualization (figure 5, [10]) was not driven by user requests, but just an experiment to find out if it could be done at all. At the Information Visualization Symposium in 2004 we got two messages back. Alfred Kobsa found the usability limited, compared to other methods [11]; on the other hand, Stuart Card showed this image in his keynote speech as an example of a nice visualization. Is this a good visualization or not?

Finally, in my own work, I found aesthetic criteria on new methods to be guiding and effective. Sometimes, each link of the chain from idea, mathematical model, algorithm, implementation to visual result is clean, simple, elegant, symmetric, etc. It is amazing how much effort is required to reach this. Developing great ideas is simple, rejection of bad ideas takes all the time.

6.3 Science

Apart from considering visualization as a technology, or as an art
21 for its own sake, we could consider visualization research as a scientific discipline. If there is something like a Science of Visualization, what should it bother about? Loosely defined, a scientific discipline should aim at a coherent set of theories, laws, and models that describe a range of phenomena, have predictive power, are grounded in observations, and that can be falsified.

If we look at the field now, many algorithms and techniques have been developed, but there are few generic concepts and theories. One reason for the lack of fundamental theories is that visualization is intrinsically complex, has many aspects, and can be approached

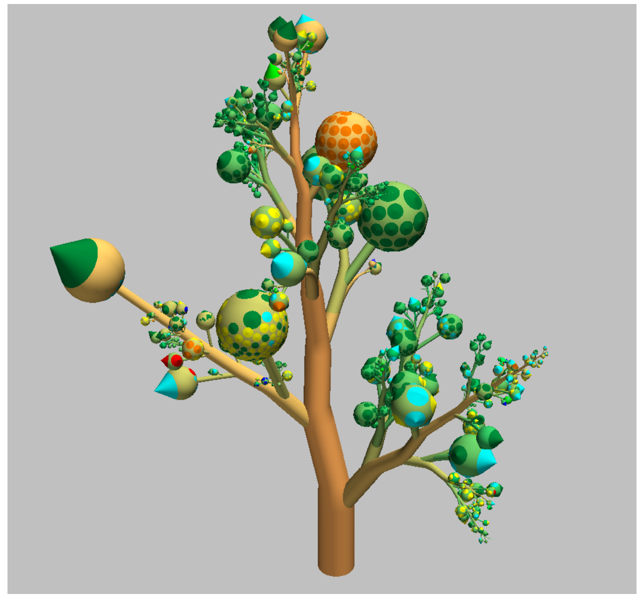

Figure 5: Botanic visualization contents of a hard disk [10, 27]. Useful or just a nice picture?

from different perspectives. In terms of the model proposed, visualization can be observed from the point of view of the data D to be visualized, the various solutions proposed (S and V), from the ΔK aimed at, i.e., the purpose or discipline for which it is applied, the images I themselves, or from aspects such as perception P or exploration E. Also, developing good visualization solutions is intrinsically a design problem, and closed form solutions for the optimalization problem "Given D find V such that ΔK is optimal" cannot be expected.

Nevertheless, we could and should aim at more generic insights, at several levels. First of all, a *descriptive* approach can be pursued further. Methods are analyzed and categorized, leading to taxonomies that show how they relate to and differ from each other. Such taxonomies span up the current solution space, and can lead to insight where new opportunities are. Some examples of good overview papers are [30, 6, 13], a great example of a taxonomy is given in [4], where a variety of different marching cube style algorithms are brought under one umbrella using computational group theory. Even if it were only because the field is still developing and overviews are quickly outdated, more work in this area should be encouraged. Taxonomies need not be confined to methods, also taxonomies on different kinds of data and especially on different types of knowledge that are relevant for end users are useful.

Secondly, *evaluation* and *validation* are important. Assessment of the effectiveness and efficiency of different methods and techniques is vital from a technological point of view (which method to use), but also as a base for more generic statements on visualization. A science of visualization should be empirical, in the sense that concrete measurements of the phenomena studied are done, which in our case concern people making and watching images that depict data. Tory and Möller [20] give a good overview of the current status of the use of human factors research in visualization, and identify areas for future research.

Thirdly, in line with the previous, we should ultimately aim at *generic results* (models, laws) that enable us to understand what goes on and to predict why certain approaches do or don't work. In the end, explanations should be based on properties of the environment of visualization, especially the end user. The value of visualization is ultimately determined by his perceptual abilities, his knowledge on the data shown, the value he assigns to various insights, and the costs he is willing to spend.

Ware's book on Information Visualization [29] is a rich source of insights on perception and how these can be used to improve visualization, Tufte gives many useful guidelines and recommendations in his books [23, 21, 22]. However, many of these are not quantitative, and also, do not explain how to handle conflicting requirements. One operational and practical criterium on guidelines is that they should allow for automated implementation, such that the user gets a good, if not optimal view on the data without costs. The early work of Mackinlay [15] on automated generation of visualizations is great, and it is surprising that the state of the art in this area does not seem to have advanced much further since then.

Finally, *methodological* issues have to be studied further. This concerns questions like how to design visualizations and how to measure and evaluate the effectiveness of various solutions. And also, how to assess the value of visualization in general.

7 Conclusion

In the preceding sections, I have tried to answer the question how the value of visualization can be assessed. As a conclusion, I think there is not a single answer, but that it depends on the point of view one adopts. One view is to consider visualization purely from a technological point of view, aiming for effectiveness and efficiency. This requires that costs and benefits are assessed. The simple model proposed enables us to get insight in various aspects of visualization, and also to understand why certain classes of methods have success and others not. Another view is to consider visualization as an art, i.e., something that is interesting enough for its own sake, and finally a view on visualization as an empiric science was discussed.

Obviously, these three different views, schematically depicted in fig. 6, are strongly related, and results from one view can stimulate work according to the other views. Finally, each view that is adopted does imply playing a different game, and if we want to win, we should play those games according their own rules: aim for provable effectiveness and efficiency, aim for elegance and beauty, and aim at generic laws with predictive power.

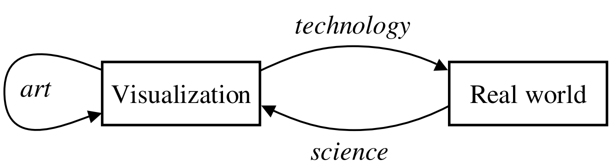

Figure 6: Views on visualization

References

[1] www.win.tue.nl/sequoiaview.

[2] www.ks.uiuc.edu/Research/vmd.

[3] Position papers NIH/NSF fall 2004 workshop on visualization research challenges, 2004. visual.nlm.nih.gov/evc/meetings/vrc2004.

[4] D.C. Banks and S. Linton. Counting cases in marching cubes: Toward a generic algorithm for producing substitopes. In *Proceedings IEEE Visualization 2003*, pages 51–58, 2003. (Best Paper award).

[5] B. Cabral and L. C. Leedom. Imaging vector fields using line integral convolution. *Computer Graphics*, 27:263–270, 1993. Proceedings SIGGRAPH'93.

[6] I. Herman, G. Melanon, and M.S. Marshall. Graph visualisation in information visualisation: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–44, 2000.

[7] B. Hibbard. Top ten visualization problems. *SIGGRAPH Computer Graphics Newsletter*, 33(2), 1999.

[8] B. Johnson and B. Shneiderman. Treemaps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of IEEE Visualization '91*, pages 284–291. IEEE Computer Society Press, 1991.

[9] C. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17, 2004.

[10] E. Kleiberg, H. van de Wetering, and J.J. van Wijk. Botanical visualization of huge hierarchies. In *Proceedings IEEE Symposium on Information Visualization (InfoVis'2001)*, pages 87–94. IEEE Computer Society Press, 2001.

[11] A. Kobsa. User experiments with tree visualization systems. In *Proceedings 2004 IEEE Symposium on Information Visualization (InfoVis 2004)*, pages 9–16, 2004.

[12] D.H. Laidlaw, R.M. Kirby, J.S. Davidson, T.S. Miller, M. da Silva, W.H. Warren, and M.J. Tarr. Quantitative comparative evaluation of 2d vector field visualization methods. In *Proceedings IEEE Visualization 2001*, pages 143–150. IEEE Computer Society Press, 2001.

[13] R.S. Laramee, H. Hauser, H. Doleisch, B. Vrolijk, F.H. Post, and Daniel Weiskopf. The state of the art in flow visualization: Dense and texture-based techniques. *Computer Graphics Forum*, 23(2):203–221, 2004.

[14] B. Lorensen. On the death of visualization, 2004. In [3].

[15] J.D. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1986.

[16] Bruce H. McCormick, Thomas A. DeFanti, and Maxine D. Brown (eds.). *Visualization in Scientific Computing*. ACM SIGGRAPH, New York, 1987.

[17] T.-M. Rhyne, B. Hibbard, C. Johnson, C. Chen, and S. Eick. Can we determine the top unresolved problems of visualization? In *Proceedings IEEE Visualization 2004*, pages 563–565, 2004.

[18] B. Shneiderman, 2004. Comment during session at InfoVis 2004.

[19] J.J. Thomas and K.A. Cook (eds.). *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE, 2005.

[20] M. Tory and T. Möller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, 2004.

[21] E.R. Tufte. *Envisioning Information*. Graphics Press, 1990.

[22] E.R. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, 1997.

[23] E.R. Tufte. *The Visual Display of Quantitative Information (2nd edition)*. Graphics Press, 2001.

[24] J.J. van Wijk. Spot noise: Texture synthesis for data visualization. *Computer Graphics*, 25:309–318, 1991. Proceedings SIGGRAPH'91.

[25] J.J. van Wijk. Image based flow visualization. *ACM Transactions on Graphics*, 21(3):745–754, 2002. Proceedings of ACM SIGGRAPH 2002.

[26] J.J. van Wijk and H. van de Wetering. Cushion treemaps. In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)*, pages 73–78. IEEE Computer Society Press, 1999.

[27] J.J. van Wijk, F. van Ham, and H.M.M. van de Wetering. Rendering hierarchical data. *Comm. ACM*, 46(9ve):257–263, September 2003.

[28] J.J. van Wijk and E. van Selow. Cluster and calendar-based visualization of time series data. In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)*, pages 4–9. IEEE Computer Society Press, 1999.

[29] C. Ware. *Information Visualization: Perception for Design. (2nd Edition)*. Morgan Kaufman, 2004.

[30] P.C. Wong and R.D. Bergeron. 30 years of multidimensional multivariate visualization. In G.M. Nielson, H.H. Hagen, and H. Mueller, editors, *Scientific Visualization – Overviews, Methodologies, and Techniques*. IEEE Computer Society Press, 1997.