Scientific Data Visualization: Can Stereoscopy Be Key?

Gunnar Sacher

CIS-TR-98-01
January 1998

Department of Computer and Information Science
University of Oregon

Scientific Data Visualization: Can Stereoscopy Be A Key?

Position Paper for the Oral Comprehensive Exam
October 31, 1997

By Gunnar Sacher, Ph.D. student

Department of Computer and Information Science
University of Oregon
Eugene, OR 97403-1202

The area of stereoscopic visualization is defined as the union of four different science areas: Visualization, Computer Graphics (i.e. stereoscopy), Psychology (i.e. depth perception), and Human-Computer Interaction. The paper concentrates on the complex integration of various perceptual cues and rendering and interaction in stereoscopic environments. It discusses human factors of stereoscopic visualization systems and argues that they must be integral part of the design of visualization tools. Furthermore, it highlights the role of the user as a central part of the visualization process.

Keywords: scientific visualization, depth perception, cue theory, stereoscopy, 3D user interfaces, 3D interaction, stereoscopic displays, stereopsis, human factors, virtual reality

0 Introduction

Scientific visualization is a relatively new field in computer science. With the development of powerful graphics computers and new viewing and display techniques it has quickly developed into a major interdisciplinary research area. It is of particular importance that the visualization community has defined the field as more than just the collection of ad hoc techniques and pretty pictures bringing a much needed scientific methodology to visualization techniques to explore and analyze an ever increasing amount of data.

This paper focuses on stereoscopic visualization. In chapter 1, we show that scientific visualization as a research methodology is both broad and complex. Utilizing our ability to see, researchers use visualization to analyze data and its underlying phenomenon to make advances in science and technology. By describing a possible framework for scientific visualization and introducing some of the important, but more general, visualization techniques we discuss the building blocks used to develop visualization tools. These tools have become a fundamental requirement in many scientific fields. We argue that their development as a combination of graphics software, powerful 3D graphical processing hardware, and new virtual reality (VR) input and output devices requires the combined effort from experts of different scientific fields. In particular, we discuss virtual reality as a powerful and creative way of interacting with computer-based systems, but also as a technology that still has limitations for achieving true 3D visualizations.

Users of visualization tools are a very active part of the visualization process. The optimal use of new technologies, such as VR, can be guaranteed only, if we understand that visualization is a user-driven process. To support this idea, we focus on stereoscopic visualization because there is a growing interest in employing stereoscopic information in visual displays. If system designers have a thorough understanding of stereoscopic cues, they can implement these cues tailored to the individual user. In chapter 2, we discuss in detail certain aspects of binocular vision and human depth perception to develop an awareness of possible perceptual interaction in stereoscopic displays. We show in chapter 3 how the geometry of binocular vision and our knowledge about depth cues can be applied to stereoscopic displays.

Utilizing stereoscopy, the design of usable and effective virtual worlds is a new challenge to developers of visualization systems. In chapter 4, we argue that, when striving to achieve the full potential of stereoscopy, we cannot separate design from human factors issues as it was possible in the past. Clearly, system limitations will decrease user acceptance of the technology. But it is also important to realize that certain characteristics of users, in addition to system limitations, influence the performance of stereoscopy in virtual worlds. Hence, human factors concerns must be part of a methodological basis for future research in scientific visualization involving VR technology, and stereoscopic displays.

Stereoscopy will be an integral part of emerging visualization techniques and technologies with an expanding variety of applications. But it is clear that more research needs to be done to maximize the efficiency of human task performance in virtual worlds created by stereoscopic displays. In chapter 5, we lay out research directions that have the goal to minimize the effort of perceiving and understanding information displayed in stereoscopic visualizations and, hence, meet better the user's need.

1 Visualization—More Than Just Pretty Pictures

1.1 Scientific Visualization—what for?

Visualization is not new to the scientific community and has often been a milepost to showing scientific progress. For many of us, the ability to visualize a problem is synonymous to understanding a problem. Minard's classic map/chart of the march of Napoleon's army in Russia in 1812/1813 and Snow's map of deaths from cholera in a part of London in 1855 are classic examples of visualizations of quantitative information that aid in the analysis of historic events (Tufte, 1983).

In 1962, at the advent of the modern computer age, Richard Hamming (Hamming, 1962) wrote: "The purpose of computing is insight, but not numbers". This emphasizes that data has to be interpreted in its context. Scientific visualization is concerned with the graphical exploration of data and information in a scientific context. In contrast to presentation graphics, which communicate information already understood, scientists want to gain better insight and understanding of data through scientific visualization. To achieve this goal, scientific visualization requires the integration of several areas of science: computer graphics, human-computer-interface methodology, image processing, system design, signal processing, data mining, perception, and cognition. Techniques from these diverse areas are used to create methods for generating and presenting large amount of numerical data containing an abundance of information that researchers want to "see". In short, scientific visualization should not be understood just as the generation of images, but as a part of a toolkit to make scientific progress.

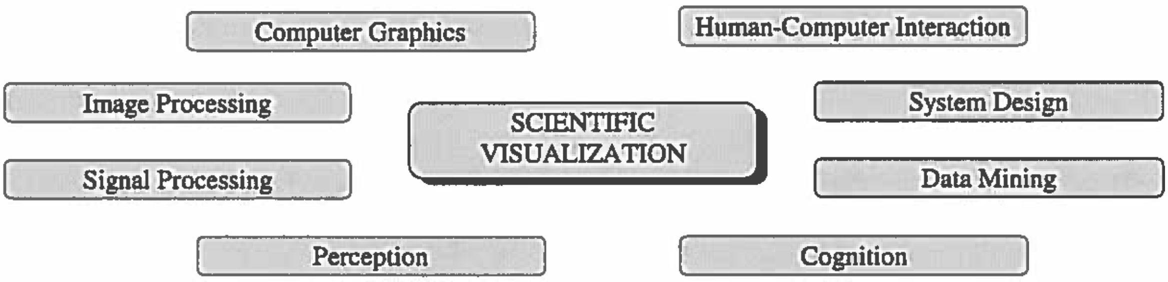


***Figure 1:** The discipline of Scientific Visualization*

Keller & Keller (1993) regarded visualization as a three-folded approach:

- identify the visualization goal,
- remove mental roadblocks,
- decide between data and phenomena.

Questions, such as 'What do I look for in this particular data set and what is important about it?', or 'What do I expect to learn and confirm?', help to identify the goal of a specific visualization and the desired result before selecting the visualization techniques. It is important to emphasize there exists a methodology in scientific visualization that involves more than one step. In particular, the exploration of data is often an iterative process, which is accompanied by interactive adjustments to the data, and searching for new relationships.

Data representation and format is certainly of concern in this process, but it is the data semantics and not its format that should determine the visualization technique used. Thinking about data only as numbers, representing information to be visualized eliminates artificial constraints such as having a certain structure or belonging to a particular application.

Keller and Keller (1993) distinguish between data-representation techniques and contextual-cue techniques. To visualize the numeric values of data, the former take data as input and deliver an image, as output (e.g. histograms, bar charts, etc.). The latter introduce special effects to an image by either adding or removing graphic elements, such as perspective, haze, and others. These carefully selected contextual cues make the data values and their relationships more readable. By visualizing the phenomena behind the data rather than just its numeric values, the meaning of the informational source is more clearly laid open.

It is no wonder then that visualization is an integral part of the process of numerical simulations. In computational sciences, we are concerned with nature and its underlying laws. This task is accomplished by proceeding through a number of steps, ranging from observing a natural event to analyzing the results of a simulation of the phenomena. Visual displays of information are often an important part to understand the processes involved; for example, weather maps illustrate aspects of weather patterns. Thus, the goal of scientific visualization is the investigation of data to gain a deeper understanding of the information source and foster new insight into underlying processes. In this process, we rely on the powerful ability of humans to visualize.

The McCormick report (McCormick et al., 1987) identified possible benefits of scientific visualizations, including the

- opportunity of increased scientific progress and collaboration, and
- an increase in scientific productivity.

Additionally, the authors recommended the development of new, useful tools to distribute them for application in different research communities. They asked for more funding for both tool users and toolmakers. However, we argue in this paper that research should not only be directed towards tool making and tool application, but also towards a better understanding of the complex issues of human factors and human-computer interfaces that are important to the discipline of visualization. Indeed, one of today's challenges consists in learning how to effectively and efficiently use the abundance of tools and resources in a comprehensive visualization process.

1.2 A framework for Scientific Visualization

Examining the design of scientific visualization systems has tended to reflect a cyclic model of the process of scientific investigation (see Watson, 1990). The focal point of the model is the users / scientists, who start the process by constructing a hypothesis, based on their knowledge in a scientific field. The next steps involve data collection, computation, viewing, and interpreting the results. The scientists will often repeat this process. This model emphasizes the usefulness of visualization for the exploration and analysis of simulation results and experimental or observation data. But it also highlights that visualization can be important to form a link between hypotheses and experiments, triggering new or refined hypotheses. What are the functional and qualitative aspects of this model for visualization systems?

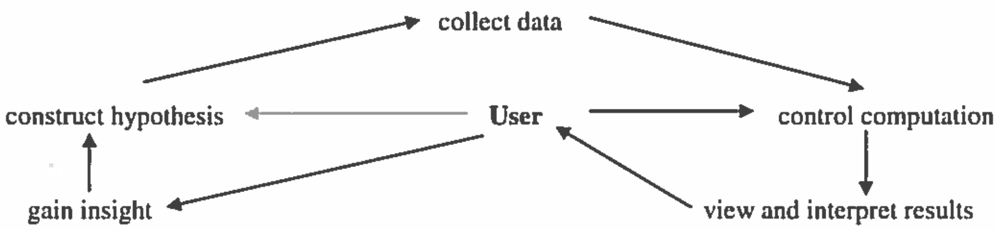


Figure 2: Watson's model of scientific investigation

A framework for scientific visualization is composed typically of modules corresponding to the *preparative*, *computational*, and *analytical* aspects of scientific investigation. These modules would support a dual dataflow. We distinguish three data paths: *input*, *output*, and *control*. The input data stream is a data flow away from the user. The output data stream is a data flow towards the user. Each module is able to respond to control data. Many visualization systems, for example Iris Explorer, support this basic visualization system design. A minimal configuration of a visualization system should include modules for data manipulation, data access, visualization techniques, basic graphics, and user interface. Other functional aspects might include extendibility, data import and export, error handling, and presentation options (Brodlie et al., 1992).

However, there are qualitative aspects that cannot be modeled or are difficult to quantify. Sometimes it is even impossible to give strict guidelines or advice. One of these aspects is *responsiveness*, that is the ability to have an appropriate response time from the visualization system. Another qualitative aspect is *appropriateness* of a system or its *generality*. Most of the visualization systems are designed for a particular area of application. Thus, in general, we want the visualization system to improve the speed with which scientific progress can be made for the application, knowing, of course, that this is often difficult to determine a priori. Sometimes, this scientific progress is not even possible without the capability to visualize data.

As difficult as these qualitative aspects are to evaluate, it is clear that they involve the interaction between the user and the visualization system. In part, this complex process has been addressed in the creation of new facilities for multimodal interaction, such as head-mounted displays (HMD) and other stereo viewing equipment, touch screens, voice, and gesture recognition. However, new techniques alone are not enough. In this paper, we argue that perceptual and cognitive issues of the human-computer interaction should be studied and integrated in visualization systems.

The interface to a visualization system should be flexible and extensible, allowing one to access different methods of data exploration and analysis. We emphasize that the main components of a visualization system—software, hardware, user, and data—have an equal weight. Unfortunately, many visualization systems are concerned primarily with the data aspect alone.

In regards to user interfaces, Brodlie et al. (1992) suggests three different aspects:

1. the interface for constructing the visualization program for a specific application,
2. the interaction with steps in the visualization process,
3. the interaction with, and exploration of model, problem, and data itself.

It is beyond the scope of this paper to discuss all three aspects in depth. In later sections, we will focus on visual interaction with a virtual world. We will concentrate on interaction and rendering in a stereoscopic environment and the complex integration of various perceptual cues, which are associated with depth, motion, scale, and orientation. Human factors of low-end VR visualization systems will be discussed in more detail. They are concerned with the comfort of the user with the system. The design of visualization systems should always include the knowledge of fields, such as perception, cognition, ergonomics, and maybe even graphical design. With their different modalities and intellect, the end-users will decide eventually about the effective use of these systems for visualization.

1.3 Visualization techniques—a brief overview

According to Brodlie et al. (1992), visualization techniques have three parts. First, an *empirical model* of the data is constructed that identifies dependent and independent variables. In a next step, the model is represented as an *abstract visualization object*. For example, assume we have a database containing 3D points of an geologically interesting area, such as a fault line. The model of this database could be depicted as a contour map or a surface view. In a last step, an engineering activity, the chosen abstract visualization object is rendered as a *sequence of images on a graphic display*. It is important to emphasize that this transformation of the data to a graphical representation can introduce errors and artifacts to the data, if the visualization technique is not carefully chosen.

Data use in a visualization can be classified based on its dimensionality. Every data entity has a range of values, typically defined over some domain of independent variables. We can express this mathematically as a function of many variables $F(X)$ where $X=(x_1, x_2, ..., x_n)$ and n is the dimension of the domain. The type of F (e.g. scalar, vector, or tensors) is used to classify visualization techniques. A further distinction can be made based on the nature of the domain because the domain dimension itself provides a sub-classification. Brodlie et al. (1992) identified three cases:

- the data set is defined *pointwise over a continuous domain* as found in terrain map
- the data set is defined over *regions of a continuous domain* as found in weather maps
- the data set is defined over an *enumerated set* as found in charts.

Other classification schemes, such as a *lattice*, have been suggested. The following summary of selected visualization techniques is based on the classification scheme introduced above.

Dimension	Point	Scalar	Vector	
			2D	3D
1	1D Scatter plot	Line graph Histogram Bar/Pie chart		
2	2D Scatter plot	Contour plot line based discrete shaded Surface View Image display Bounded Region-Plot (value associated with region is coded by means of color or shading) 2D Histogram 2D Bar chart	2D Arrows 2D Streamlines 2D Particle Tracks (static polylines indicating direction of flow) 2D Field Topology (streamlines are drawn from critical points, i.e. saddle points)	3D Arrows in plane (point into or out of the 2d display surface)
3	3D Scatter plot	Isosurfaces Basket Weave (extension of isosurfaces) Volume Rendering ray casting		3D Arrows in Volume (use additional depth cues)
n	n-dimensional Scatter Plot			

Table 1: visualization techniques (based on Brodlie et al., 1992, p.43)

The summarized visualization techniques are solely based on spatial components and assume a "plotting" model. There are other representations that are different from the ones described above. For example, animation techniques introduce a time component t; $F(x_1, x_2, \ldots, x_n; t)$. They are useful for enhancing many types of visualizations. By providing a sequence of images for the data analysis, they allow the user to gain more insight into the phenomenon underlying the data that is continuous over time. This is often crucial for the understanding of complex data and scenes.

The application of a chosen visualization technique will result in a graphical display. The ability to interact with the data is another important component of a visualization system. This interaction might be used to post-process data or to steer the generation of new data. Besides selecting areas of interest (e.g. for data probing, or annotating images), often we want to change general viewing properties. Interaction techniques that accomplish this include panning, zooming, and changing view point and camera positions. For many applications it is essential to allow an interactive control of parameters. These parameters might not only be used to manipulate the display but also to control the data generation. Difficulties in the perception of visualizations can be often overcome or at least eased by giving the user control to data associated with a certain image, such as color maps (see 2.2).

1.4 Visualization software, tools and environments

This section provides an overview of visualization software products that support the visual exploration of data. In the early days of visualization, it was rather painful for the scientists to produce visualizations beyond plots and simple graphics because of the graphics programming skills required. Over the past years, many visualization systems have been developed that require fewer programming skills and allow users to produce their own visualizations. In general, the visualization software tries to match the scientist's needs and requirements. These requirements include the integration of visualization and computation, the expectation of a highly usable and productive environment, the ability to manage large and multidimensional data sets, and the ability to interact in real-time (Earnshaw, 1994).

Meeting these requirements can be difficult. Currently, we can distinguish four major visualization software categories (see (Brodlie et al., 1992) or (Earnshaw & Wiseman, 1992) for more detailed descriptions and references). *Graphics libraries and presentation packages* are the more traditional approach to visualization. Graphics functionality is provided in software or by interfacing directly to the hardware. Programmers provide almost all the components their application is composed of: the main program, the user interface, the data handling, and the geometry mapping. A representative of this approach is OpenGL (Silicon Graphics). Programs using graphics libraries require expertise to write and the applications are often difficult to maintain and modify. Presentation packages, such as Harward Graphics or MatLab, provide user interface functions but still require a lot of effort to create the visualization. *Turnkey visualization systems*, such as PHIGS PLUS for geometry data or the Data Visualizer (Wavefront) for scientific data, provide the user with data and instructions to the main program. Offering a usually attractive interface, these systems take care of the rendering as well but typically they cannot be modified. *Higher-level application builders* provide turnkey solutions for individual parts of the program as modules. Although no programming is required, these modules can be replaced or rewritten, providing flexibility and extensibility. In addition, users can add their own modules, which have to comply with data input/output conventions, to customize their final solution. When building an application, a network of modules is created. Examples of application builders are IRIS Explorer (Silicon Graphics), IBM's Data Explorer, and Khoros (University of New Mexico). The weak points of these systems are their interaction techniques and the handling of large data sets.

Earnshaw (1994) suggested a new generation of 3D data visualization technology. He proposed a *steering model*, which is a direct interactive model, that enables the user to analyze large data sets interactively. It tries to overcome some of the weak points inherent in application builders by allowing the user to explore very large data sets and interact with them directly. This is essential for a better understanding of the details in large data sets and enables visualization tools to become more user-friendly. The key to Earnshaw's model is that all operations on data, such as filtering, mapping, and rendering, are performed only in a sub-region of the data, selected by a local query. In application builders, simulations must be run first, before the collected data can be visualized. Contrasting this batch-processing approach, the steering model integrates the simulation with the visualization. The data are visualized directly as they are generated. The system users then are able to interactively steer the computation to regions of their interest in the simulation.

Using software described above, developing scientific visualization systems often has been proven to be difficult. The implemented visualization techniques are too slow or the systems do not provide sufficient performance in order to respond interactively and quickly. There is obviously a need for real-time visualization systems that ideally enable the user to explore data sets in a dynamic process often by means of virtual reality tools and environments. But real-time performance is constrained by two issues (Bryson, 1996):

1. the system must provide feedback to the user in less than 0.1 sec to ensure a continuous user input,
2. the animation rate must be equal or greater 10 frames per sec.

These requirements often conflict. Therefore, system developers have to compromise speed and accuracy, speed and limited data access, and/or speed and quality of rendering. The user and not the system developer has to make these compromises. It is the user who has run-time, interactive control

- over parameters and algorithms that determine the performance of the computation,

- over data representations that depend on the data to be visualized and the desired method of visualization, over data management, such as data sampling and compression strategies, and
- over the form of graphical representations, such as low-level versus high-level graphics primitives.

There are many limitations to the approaches above. Most of the systems are not optimized to handle large programs and large data sets. Users wanting to interact with their systems, extend them, or maintain them, still need a lot of expertise. Furthermore, the data-flow driven, visual programming style is not always natural to the user. Newer visualization technology can help to address these problems. The following key requirements attempt to define the next-generation data visualization systems (Earnshaw, 1994):

- specific high performance graphics hardware must support 3D rendering because performance is often sacrificed for standardization in standard graphics software systems,
- independent of size and complexity of the data, direct interactive manipulation of data must be available,
- it should be possible for the user to integrate different visualization techniques into one image
- as a prerequisite for steering and control systems, the user must be able to semantically interact with the data assuring data probing in the application data domain,
- visualization is needed in different contexts,
- the intelligent filtering of the data should support the handling of very large data sets allowing only a subset of the data to be rendered,
- real-time visualizations are needed,
- new presentation techniques need to be added to visualization systems, e.g. videocopy, stereoscopic overheads,
- based on perception, mental models and interaction, visualization should be increasingly user driven.

Clearly, the visualization research problem as a whole is both very broad and complex, and, hence, this paper does not attempt to initiate a detailed discussion of the research space outlined above. However, it is interesting to regard visualization as a user-driven process. New techniques then are really part of a much larger visualization environment involving the user. In the following chapters, we concentrate on a narrow, very specific problem—stereoscopic scientific visualization (see Fig. 3)—where some of the above listed requirements arise. In particular, the utilization of depth cues is discussed as an important factor that, if correctly applied, can enhance information in computer displays utilizing new visualization techniques (e.g. stereoscopic display techniques). We focus on how we perceive depth in natural and virtual environments and how we can create stereoscopic displays in which some functional aspects of our visual system are crudely mimicked. Furthermore, using stereoscopic display techniques as an example, we show how issues of human perception and user-driven visualization are closely related to each other.

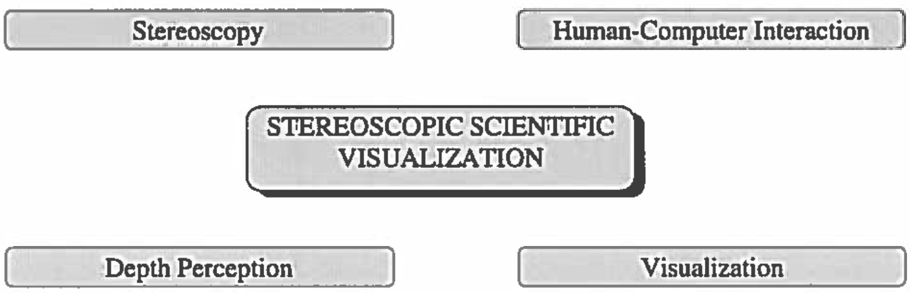


Figure 3: *Stereoscopic Visualization as discussed in this paper*

We hope to demonstrate in the next chapters why we think it is useful to invest energy and resources into the development of stereoscopic applications. We believe that for many scientific problems the use of stereoscopy in scientific visualization is a natural progression from today's two-dimensional and three-dimensional flat windowing systems. Our goal is to show that stereoscopic visualizations provide an unambiguous display of complex three-dimensional structures by using a rich set of depth cues. In addition, these displays allow natural and rapid explorations of the volume containing the data.

2 Depth—Perceiving the Third Dimension

2.1 Perceiving three dimensions

Although a lot of research has been done over the past decades to gain more insight into the very complex processes of human vision, the very nature of seeing is still a mystery to scientists. Clearly, the estimation of depth, which is one of the most involved processes in vision, was and is crucial for our survival. Research on the mechanisms of human depth perception has been guided by two main theoretical approaches. A *constructivist approach*, also called *cue theory*, has emphasized that it is possible to isolate a number of information sources that the perceptual system uses to estimate depth. Depth perception and the closely related operations of size perception are considered to arise from the computation and synthesis of these many depth cues. An alternative analysis of human depth perception, termed the *ecological approach*, has emphasized that depth relationships can be perceived directly from invariant information in an optic array. It is important for developers of stereoscopic visualizations to understand these two approaches because they make extensive use of these principles of human perception when creating displays.

2.1.1 The cue theory

The cue theory assumes that vision involves apprehending many sources of information about objects in the environment and about how our visual system interacts with them. These many cues are interpreted by our perceptual system, allowing us to make inferences about depth in a scene. According to this theory, the connection between these visual cues and real-world depth is learned by experience. . It is known and often emphasized (MacAdam, 1954; Ramachandran, 1986) that only the synergy of different depth cues lets us experience depth as we do in our everyday life. Several groups of depth cues are distinguished. *Physiological depth cues* are related to the anatomy of the eyes and the physiological processes involved in vision. They include *oculomotor cues* and *binocular disparity*. *Psychological cues*, such as *motion-based cues* and *pictorial cues*, form a second group. These cues arise from the information available on the two-dimensional surface of the retina.

2.1.1.1The physiological depth cues

Oculomotor cues are *convergence, divergence,* and *accommodation*. They are defined by our ability to sense the position of our eyes and tension in our eye muscles. Convergence occurs when eye muscles cause the eyes to look inward focusing at something very close. In contrast, divergence occurs when we look at something far away causing the eyes to move outward. Accommodation is the flattening and bulging of the lens to focus on objects. Oculomotor cues can serve as depth cues when the position of the eyes and the shape of the lens are correlated with the distance of objects, which we are focusing at. In the literature, there was confusion whether this is true. However, research showed that these cues are effective at a distance closer than about 150 - 300 cm from the observer.

Binocular disparity is the major depth cue that depends on both eyes. Because of the two different locations of our eyes, we perceive two slightly different views of our surrounding environment. The physicist Charles Wheatstone (1838) was one of the first who built a stereoscope that used two slightly different images for the left and right eyes to produce an illusion of depth. Therefore, the stereoscope takes advantage of binocular disparity.

Corresponding retinal points are points on each retina that overlap if one retina is slid over the other (A and A', B and B', C and C' in Fig. 4). Every point on one retina has its corresponding point on the other. Objects falling on the *horoptor,* that is an imaginary surface that passes through the fixation point, create retinal images that fall on corresponding points on the two retinas. Retinal images of objects that are not located on the horoptor fall on *noncorresponding,* or *disparate points* (A and C' in Fig. 4). The distance between two noncorresponding points is called *degree of disparity,* which is the distance that one of the retinal images must be moved so that both retinal images fall on corresponding points. The degree of binocular disparity increases with the distance of the object from the horoptor.

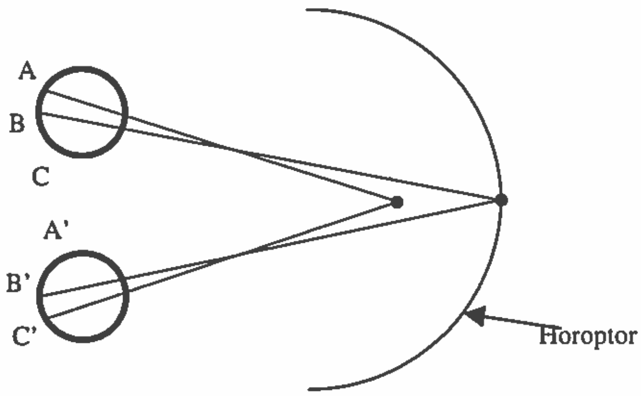


Figure 4: Concept of corresponding retinal points

Disparity in front of the horoptor is called *crossed* or *negative* and behind the horoptor is uncrossed or *positive*. The concept of disparity and horoptor is used in several applications. Developers of stereo displays call the difference between corresponding points in the left and right stereo image *parallax*, which can be directly measured in the image plane. As we explain in Chapter 3, parallax produces retinal disparity. We distinguish *horizontal* and *vertical* parallax. The former corresponds to retinal disparity and contains the depth information in the stereo image. The four types of horizontal parallax are summarized in Table 2 on the next page. Divergent horizontal and vertical parallax are not found in natural scenes and often cause pain or eye strain, when observed in stereo displays.

Let us return to the original definition of disparity. Disparity determines depth correctly only up to a scale factor. The retinal images of a 3D object become less disparate with increasing viewing distance. They are equal for an object infinitely far away. Therefore, a given disparity value corresponds to a larger distance from the horoptor for a far away object than for objects that are close to the horoptor. Buelthoff, Fahle, and Wegmann (1991) report that perceived depth depends not only on disparity but also on the disparity gradient. They discovered that when two stimuli approach each other laterally, the perceived difference in depth between them decreases although the objects stay in the same fronto-parallel plane because their disparity difference is kept constant. This means, with an increasing disparity gradient *G*[1], the two stimuli seem to approach each other not only laterally but also in depth, that is one object seems to approach the horopter from in front of the screen and the other from behind the screen. Buelthoff et al. demonstrate that our depth impression depends not only on the disparity but also on the disparity gradient between objects. Their data show clearly that this effect is more evident for larger disparities, that is the larger the disparity the more noticeable is a decrease in subjective depth when the disparity gradient increases. Interestingly, the strength of this effect depends on the shape of the stimuli. It is more pronounced for line stimuli than for small symbols and point stimuli and almost insignificant for large symbols. Furthermore, the scientists could show that the described effect of a decrease in perceived depth with increasing disparity gradients is more obvious for stimuli with an horizontal orientation (0°) rather than an oblique (45°), or vertical (90°) orientation. Keeping the disparity gradient constant by increasing the absolute disparity between two objects $|(d_L - d_R)|$ and the distance between the stimuli also decreases the proportion of perceived depth difference to the disparity difference of the displayed objects. The experimental results lead to the conclusion "that we are more correct in our depth estimates for steep gradients in depth when the Euclidean distance between the stimuli is short" (Buelthoff et al., 1991, p.150), that is when the absolute disparity between the stimuli is small. This fact is very important for the design of stereoscopic images containing

[1] G is defined by the ratio of the absolute disparity between two objects $|(d_L - d_R)|$ and the binocular separation S_{bin}, that is $G = |(d_L - d_R)| / S_{bin}$ (Burt & Julesz, 1980)

multiple objects. Buelthoff et al. also suggest to use different or large symbols and to avoid symbol arrangements of horizontal orientation. Opposite contrasts of the stimuli can further ease the problem.

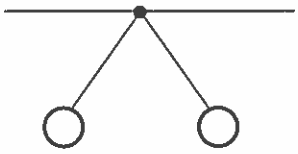
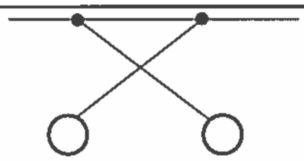
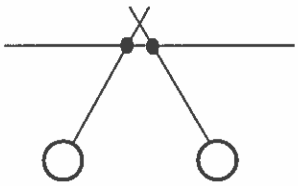
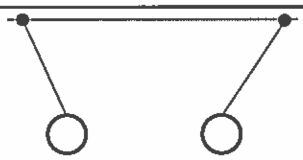
Zero horizontal parallax		objects are seen on the surface of the screen because the left and the right eye (circles) converge at the plane of the screen (dashed line), that is the lines of sight cross at the display screen; image points have zero parallax setting (zps)
Negative horizontal parallax		objects are floating in front of the screen because lines of sight converge in front of the screen surface; parallax points are crossed, or negative
Positive horizontal parallax		objects are within screen space, that is objects are floating behind the surface of the screen because lines of sight converge behind the screen surface; parallax points are uncrossed, or positive
Divergent horizontal parallax		positive horizontal parallax is larger than interpupillary distance, that is the distance between the two eyes; unusual muscular effort to fuse images can cause discomfort

Table 2: Horizontal parallax

2.1.1.2 The psychological depth cues

In this section, we examine psychological depth cues, such as pictorial cues. Those cues are derived from the 2D image on the retina. Pictorial cues are a powerful group of depth cues that belong to the group of monocular cues. They are cues that can be represented in a picture and include *light and shade, occlusion* (or *interposition*), *size* and *height in the field of view, atmospheric* and *linear perspective,* and *relative* and *familiar size.* These monocular cues are part of every electronic display and as such provide the basis for depth perception in those visual displays.

Light and shade are very basic depth cues. From experience, we know that bright objects appear closer than darker ones. Adding shade to drawn objects makes those objects look more solid and their shape more realistic looking. An object that casts shadows often appears to be attached to a surface.

Similarly, occlusion (or interposition) is a very obvious cue. If one object covers another, we perceive the covering object usually in front of the covered object. This indicates relative depth because one object is seen closer than the other.

Size in the field of view and height in the field of view are two other important depth cues. Assume two objects of the same type at different distances from the observer to explain the former cue. The object that takes up a large part of the field of view is closer to the observer than the object that takes up only a small part of the field of view. Ames demonstrated that larger size causes an object to appear closer (Ittelson, 1952). In an experiment which he had designed, observers viewed illuminated balloons in a darkened room. They reported that balloons that expanded in their volume appeared to be moving closer. Assume again two objects of the same type below a horizon to illustrate the height in the field of view cue. The object that is higher in the visual field will be seen usually as farther away. Objects above the horizon line seem to be more distant when they are in the lower field of view.

Atmospheric perspective causes us to see objects in the distance as more blurred and bluer. This perception is caused by particles in the atmosphere and the amount of air we have to look through. Therefore, thick fog or haze make objects appear to be more distant. Peaks of mountains, seen in the distance, will often show a bluish color because the red part of the light gets scattered by the atmosphere.

Artists have always been concerned about how to recreate a three-dimensional environment on a two-dimensional piece of canvas. In 1435, Alberti described a technique that has been called Alberti's window and allows anyone to draw in perspective. The principles of linear perspective are based on the fact that parallel lines converge towards each other in the distance; greater convergence of parallel lines indicates greater distance.

The concept of relative size takes the retinal image size of an object into account. When comparing two same sized objects we have learned that the larger object is closer to us than the smaller one. Our knowledge of the actual size of an object, that is its familiar size, sometimes influences the distance we perceive the object at (Predebon, 1993). Predebon showed that depth estimates are greater for a small-sized familiar stimulus than for the same normal-sized stimulus. Estimates of depth for a stimulus of unfamiliar size were unaffected by a size change that was equivalent to the change in size for the familiar stimulus. This *familiar size* cue is most effective when other depth cues are absent.

Deletion and *accretion*, and *motion parallax* are motion-based psychological depth cues. Assume two objects at different depths from the observer. Head movement perpendicular to the line of sight to the objects gives the impression that the objects move relative to each other. Moving in one direction covers, or deletes, the distant object more and more. Moving in the other direction uncovers, or accretes, the distant object. Deletion and accretion are related to occlusion and the third motion-based depth cue: motion parallax. Motion parallax is based on differences in speed for near and far objects, as an observer moves. Near objects move more rapidly than farther objects. Using this cue, we perceive depths of objects by judging their speed as we move. This effect occurs because moving near objects cover a larger distance across the retina than moving farther objects, and therefore move more rapidly across the retina as the observer moves. Motion parallax is exploited by flight simulator panoramic display systems. When no movement is simulated, the user uses pictorial cues to obtain depth information. However, when the simulated airplane moves, the user experiences an enhanced depth impression.

As already mentioned, in the case of depth cue conflicts, our mind will try to solve the conflicts by either favoring one of the cues or giving us a false depth impression. If none of the solutions is chosen, we might see *diplopic* images. Viewing scenes with contradicting depth cues often can cause discomfort. MacAdam (1954) points out that if either perspective or disparity deviates from our normal visual experience a compensation of one by the other is not possible. This fact is a very important guideline for the generation of stereo images because it makes heavy use of these two cues.

2.1.2 Gibson's ecological approach to depth perception

Gibson (1979) proposed another theory for depth perception. His ecological approach stands in contrast to the cue-based constructivist approach described above. The key to Gibson's depth perception theory is the *optic array* and its sources of information. In contrast to depth cues, the informational sources in the optic array contain invariant information that is information that does not change as the observer moves and is perceived directly without involving mental processes. Gibson's theory can be summarized in three basic principles:

1. Movement should be emphasized because we are steadily moving when perceiving the environment.
2. Rather than studying the informational content of a static retinal image, the informational content in the optic array—a structured pattern of light received by the eyes—should be evaluated.
3. Informational processing is unnecessary because there is enough explicit information in the optical array. Gibson calls this *direct perception.*

Among the informational sources in the optic array are *texture gradients, flow patterns*, and the *horizon ratio*. Texture gradients are patterns of textured surfaces. The size of the pattern elements gives us, the observer, an impression of depth, that is the pattern element size decreases with increasing distance from us. Furthermore, texture gradients contain information about surface orientation and provide a basis for size constancy. In the depth cue approach, size constancy is the result of a calculation that is not necessary in the ecological approach. An example for invariant information in texture gradients is the spacing between the texture elements.

Using texture and size cues, Beverly and Regan (1982) studied the effectiveness of these cues as stimuli for motion in depth. The observation that an object moving towards the eye of an observer causes two changes to the retinal image of this object—size and coarseness of the surface texture increase—was the bases for a series of experiments. Object size and texture of stimuli were expanded or contracted simultaneously, or were changed so that one or the other expands rather than contracts. This research has important implications for the design of visualizations because they were able to show that changing object size is a very important cue to perceive motion in depth. A combination of depth cues, such as object size and texture gradients, did not necessarily produce a better depth display for the observer. In general, the single cues had to be at least consistent with each other. In the case of monocularly viewed, two-dimensional motion-in-depth displays, the addition of texture to a stimulus without texture did not increase but reduce the effectiveness of this stimulus.

Flow patterns provide similar information as the depth cue of motion parallax. Rather than emphasizing the motion of a single object, flow patterns provide information on the flow of the whole visual field. The speed of flow is higher in the foreground than in the background.

The horizon ratio is another source to perceive size information. It is defined by the ratio of the object proportion above and below the horizon. The *horizon ratio principle* states that the observer's eye level is equal to a height that is defined by the intersection of a distant object and the horizon line. Furthermore, if the proportions of two grounded objects above and below the horizon are the same then we can assume that these objects have the same height. The horizon ratio of objects in a scene is invariant, that is it does not depend on the position of an observer in a scene. Hence, without further processing, we directly perceive size information that can be used to make depth judgments (see also 2.1.4).

2.1.3 Cues versus optic array—a concluding remark

The question is not whether the constructivist approach or the ecological approach is correct. Both views contribute to our understanding of depth perception in that both depth cues and invariant information probably contribute to our perception of depth and size. They support the idea that the visual system combines information from different sources to produce a coherent 3D percept. Table 3 summarizes some of the cues that are important for the generation of stereo images.

aerial perspective	sharp objects appear close in distance, hazy objects appear distant
light and shade	bright objects appear close in distance, dim objects appear distant, shading gives objects the look of being solid or rounded
linear perspective	parallel lines converge in distance
motion parallax	fast moving objects appear close in distance, slow moving objects appear distant
occlusion	close objects overlap more distant objects
relative image size	large objects appear close in distance, small objects appear distant, memory helps to make judgment about the distance of familiar objects
textural gradient	detailed objects appear close in distance, objects with no detail appear distant

Table 3: psychological depth cues

One informational source alone is often not enough to determine the depth percept we experience. Braunstein, Anderson, Rouse, and Tittle (1986) looked at the effects of combining occlusion, a monocular depth cue, and binocular disparity information on the recovery of depth order. Their results indicate that in the case of conflicting monocular and binocular cues the monocular information is able to override the binocular cue. Sacher, Hayes, Thornton, and Sereno (1997) demonstrated depth reversals using size and binocular disparity cues in an orthographic display environment. They were also able to show that manipulating the size information could weaken the overriding effect. This shows that the interpretation of pattern in 3D space is an extremely complex process taking into account monocular and binocular information as a whole. In the case of cue conflicts, our spatial perception changes in a natural way that is steered by the sensory input of the most powerful cue. Looking at two objects in depth, our mind seems to have no knowledge about absolute values of binocular disparity values when it tries to interpret the scene. Only the difference between binocular disparity values in addition to other determining factors helps us to select one out of many interpretations.

2.1.4 The perception of size and shape

The perception of size is tightly interconnected with depth perception. This section summarizes some of the important ideas about this relationship. The size of an object determines its visual angle, that is the angle between two lines starting at our eye and ending at both ends of the object. The visual angle is defined relative to the observer. It changes as the distance between observer and object changes. Furthermore, the smaller an object, the smaller the visual angle. A small visual angle produces a small retinal image. Two objects will have the same retinal size if their visual angles are the same. Changing the distance to a familiar or unfamiliar object changes its visual angle but does not change our perception of its size. The perceived size remains constant regardless of the changing retinal image. This fact finds its expression in the *law of size constancy* and suggests that our perception of size is not solely depending on the retinal image size. Halway and Boring (1941) showed in one of their experiments that size constancy depends on the availability of depth information. A *constancy scaling mechanism* introduced by Gregory (1966) gives an explanation for the connection between size and depth perception. He proposes that the distance of an object is added to the information available in the retinal image. Our everyday life is not dominated by perceiving an object's visual angle but by using a size-distance scaling mechanism due to the abundance of depth cues that exist in our surrounding world. However, many size illusions are caused by erroneous depth information, for example the Ames room (Ittelson, 1952).

Binocular disparity information is one important component of processing shape of objects in a three-dimensional environment. Johnston (1991) shed some light on the effectiveness of the disparity cue in defining three-dimensional shape. In general, the study showed that depth perception is poor and often not veridical in an impoverished stereoscopic viewing environment. Even using surfaces rich in disparity information did not guarantee a correct calculation of the scaling distance. One possible answer to the question why depth information was used incorrectly might be hidden in the process of how the visual system combines different cues that specify depth and distance. In Johnston's experiments, as is the case with many stereoscopic display environments, only a subset of depth cues was present in the stereograms of squashed, circular, or stretched half-cylinders. The collected experimental data showed different systematic shape distortions for different viewing distances. For the longest viewing distance, the depth of a half-cylinder, which was viewed as apparently circular by the participants, is consistently larger than the half-height. Therefore, a physically circular half-cylinder is perceived as squashed elliptical cylinder. This means that depth of the display object is under-estimated. At an intermediate viewing distance, the presented half-cylinder are perceived as what they are squashed, circular, or stretched. At a close viewing distance, depth was persistently over-estimated, meaning that a decreased depth value is necessary for the participant to perceive a circular cylinder and therefore, a physically circular half-cylinder is perceived as stretched. We should keep these findings in mind for the design of stereoscopic visualization software. If the user is able to make adjustments to the viewing environment, such as the distance from the eye to the plane of zero parallax, certain adjustments have to be made within the application.

2.2 Chromostereopsis

In this section, we briefly introduce a depth illusion, called *chromostereopsis,* which can be created when looking at differently colored patches, which are located on one surface plane. This effect relates only to the wavelength of the color. When creating colored stereoscopic displays, it is important to understand the interaction between chromostereopsis (perceived depth based on hues) and stereopsis (perceived depth based on binocular disparity) because hues are able to falsely alter the intended depth to be perceived. Hence, chromostereopsis could act as an interference factor with accurate stereoscopic depth perception.

Vos (1960) laid the theoretical framework for chromostereopsis by looking at interactions of the natural binocular viewing process and chromostereopsis. He came up with two main theories: the Bruecke-Einthoven theory and the Stiles-Crawford theory. The former is based on the divergence between the optic and visual axes of the eye (Fig. 5). The two axes do not coincide because the fovea typically lies on the temporal side of the optical axis. This discrepancy is the cause for the dispersion of light as a function of wavelength, that is shorter wavelengths, in comparison to longer wavelengths, are reflected to relatively nasal positions. Hence, coplanar red and blue stimuli stimulate noncorresponding points, so that the red stimulus appears nearer than the blue one (red advancing bias). But some observers report an opposite effect that is the blue stimulus appears closer than the red one (blue advancing bias). Vos (1960) proposes that this conflict in the data can be resolved by the Stiles-Crowford effect. He suggests that the orientation of the foveal cones in relation to the pupil tends to counteract the dispersion of light on the retina, that is the cones are most sensitive to light that enters the human eye along an axis that is on the opposite side of the visual axis from the optical axis. The magnitude of this effect is determined by the size of the pupil and determines the direction and the magnitude of chromostereopsis.

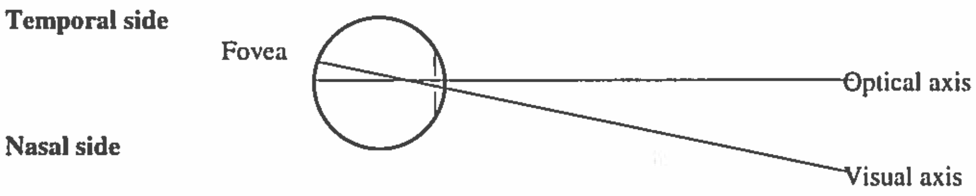


Figure 5: Exaggerated drawing of the axes of the eye

We return to the subject of chromostereopsis in Section 4.1.3 discussing performance issues in stereoscopic environments.

2.3 An important step—the step from perception to visualization

How can we integrate knowledge from the area of perception into the visualization process to improve the correct perception of visualizations? Clearly, human perception plays an important part in applications involving stereoscopy, and it is important to realize the richness and complexity of the human visual perceptual system that can be utilized in interpreting and understanding data. Hence, a better understanding of visual information processing would be expected to improve the visualization process and its results. The challenge is in being able to select the best perceptual and best sensory modalities. As noted by Davis, Corso, Barfield, Eggleston, Ellis, Ribarsky, and Wickens (1994), for specific tasks, we must know what perceptual information is critical, what simplifications are acceptable in the visualization, to which sensory modality the information should be displayed, when and if inputs to one modality can substitute or augment those of another modality, and how to coordinate or mesh those perceptual inputs from different modalities. The key viewpoint is that the human observer of visualizations is not a static object, but rather an active participant.

The integration of perception and visualization emphasizes the statement that "visualization is much more than a method of computing (Gershon, 1994)". Indeed, Brodlie states that "we not only believe what we see: to some extent we see what we believe (Brodlie, 92, p. 75)." So it seems reasonable that the visualization designer should always ask what kind of 3D object does the user see when it is displayed on a two-dimensional screen. Additional information, such as context, perspective, lighting and shading, movement, and stereo views can greatly improve the perception of three-dimensional objects (see 4.1).

Gershon (1994) focuses on three general ways to improve the visualization process:

1. Enhancing the visibility of information in the displayed data by choosing, for example, a suitable color scale and an appropriate brightness between adjacent regions in the image.
2. Increasing the faithfulness of visual representations by applying knowledge of the perceptual process.
3. Making the process of observing the visual display faster and more effortlessly by increasing the magnitude of the preconscious part of visual information processing.

Using examples from color perception, he recommends using a color scale based on brightness contrast rather than on hue contrast. Some visualizations make heavy use of pseudocolor to represent values of data points using a hue contrast based color scale. Using this scheme a data value or a range of values is associated with a color of the rainbow color scale, that is the mapping from data to color is defined by the association of a data point or a range of data to a position of a color in the color spectrum. This scheme sometimes hides the shape of objects. Therefore, a brightness contrast-based color scheme is recommended. In this scheme, a higher data value asks for a brighter color, where brightness means perceived brightness and not physical value of the brightness. Such a scheme not only helps to discriminate the data values better but also enhances the overall shape of objects in the image. It is important to note when using color in visualizations that we need to be aware of a perceptual phenomenon, called irradiation, when choosing an object and background colors. Based on the fact that we perceive a bright object on a dark background as bigger than we perceive the same but darker object on a bright background, a scaling process according to the relation of object and background color would be desirable. Taking into account that perception is an individual experience, this is difficult to implement. Brightness is also an important depth cue. Enhancing the brightness contrast of an object with its adjacent regions improves depth perception. A computer graphics technique for wire frames and solid objects, called *depth cuing*, makes use of it. Depth cuing reduces the brightness of an object in proportion to the distance from the viewer and is used to simulate the cue of aerial perspective and the cue of light and shade. In general, this stresses the need to choose a color scale in accordance to the data we want to display and the need of being aware of possible side effects.

Assuming that there is a preconscious (automatic and rapid) and a conscious (not automatic and requiring focused attention) stage in analyzing data (Treisman, 1987; Friedhoff & Benzon, 1989), supporting the preconscious stage of the visualization should speed up the conscious stage. This can be done by improving the visibility of information embedded in the visual display, for example by animation of the displayed object, or by changing the orientation of the object. Livingstone (1988) points out that the sensitivity of the visual system to "color contrast, movement, and stereopsis suggests that hard-to-see objects can be made more visible by introducing movement (by moving the object or the observer) or stereopsis (by simultaneously viewing two images of the same scene taken from different positions)".

In this chapter, we introduced depth cues and related issues that are the basis for the perception of depth in visual displays. Most of these clues are also part of every planar electronic display supporting the development of three-dimensional views. Stereoscopic displays differ from planar displays by incorporating parallax information. In the next chapter, we discuss the technical and graphical aspects for the composition of stereoscopic computer displays as a first steppingstone to high-end VR systems and the requirement of a natural, inherently three-dimensional interface. This discussion takes into consideration facts about depth perception highlighted above.

3 Stereoscopy—Visualizing the Third Dimension

Scientists were always curious about reproducing the ability to perceive depth through two eyes using special mechanical devices. For about 150 years, 3D displays have been more or less a curiosity, first, for the rich and famous, later, for the average person, who was looking for new thrills in the entertainment industry. Lately, since the early 90's, with advances in 3D viewing hardware and imaging software, it has become increasingly popular in engineering and scientific fields such as mechanical engineering, computational fluid dynamics, molecular modeling, medical imaging, industrial design, and flight simulation.

3.1 From Wheatstone's stereoscope to Lipton's stereo glasses and the Cave

In this section, a brief review of the evolution of stereoscopy is given. For a more detailed discussion, the reader is referred to one of the several publications of Lenny Lipton, one of the pioneers of modern stereoscopy (e.g. Lipton, 90 and Lipton, 91a).

In 1833, Sir Charles Wheatstone (Wheatstone, 1838) invented the mirror *stereoscope* based on the principle of stereopsis. He drew two images of a block, one for the left eye and one for the right eye. The two images showed slightly different views of the same object because they were horizontally shifted. When these two images were sent separately to each eye of an observer via two mirrors, the observer saw one "solid" appearing object in depth. Although people like Euclid and da Vinci were aware that each of our eyes sees a different image of the world, it was Wheatstone's stereoscope that demonstrated a special sense for depth, *stereopsis*, that is based on retinal disparity. Other scientists, like Kepler, had wondered why we do not see diplopic images of the world. Wheatstone could demonstrate that our visual system fuses two planar retinal images into one so that we can see one solid object.

In 1858, inspired by Wheatstone's stereoscope, J.C. d'Almeida developed another technique for viewing stereoscopic images, which is also still in use today. The *anaglyph* method uses red and green, or red and blue filters placed over the lenses of two projectors. The projected images were superimposed on the screen. People in the audience looked through glasses having a red filter for one eye and a green filter for the other, so that only one of the screen images is passed to each eye. Disparities captured in the anaglyph, matching those of the image pair seen by our eyes, make sure that the brain perceives the same information that it would perceive if an actual 3D object were presented. In the 1920s and 30s, major motion picture studios, such as MGM, used this technique to produce several short movies. Anaglyphs were printed on a single roll of film, precluding other colors than green and red. People complained that the special spectacles caused eyestrain because the used projection method put a lot of stress on the eyes.

At the beginning of this century, because of improvements in photomechanical reproduction, new viewing devices were developed to please the desires for new forms of mass entertainment. Movies became more popular and accessible for a broad audience. The inventor of the Hammond organ, Laurens Hammond, was also the inventor of one of the most interesting viewing systems for stereoscopic movies called the Teleview system. Presumably, the system was the first and only commercially used stereoscopic frame sequential motion picture system. It used two electrically synchronized 35-mm projectors, one showing the images shot by a camera for the right eye and one for the left eye. In front of each projector lens, two shutters were placed, which were synchronized out-of-phase. Visitors in the movie theater used a spinning mechanical shutter that was mounted in front of them to watch the movie. The shutters of the viewing device were synchronized in phase with the shutters of the projectors. This guaranteed that the appropriate image was displayed to each eye. When the left eye image was projected to the screen, the projector lens for the right image was blocked. In the same moment the position of the spinning shutter made sure that the left eye was uncovered to view the image. Displaying the interrupted projected images at a high enough rate, patrons perceived a stereoscopically fused, flicker free movie.

Another important 3D display technique was shown to an audience for the first time at the New York World's Fair in 1939. The *polarized-light projection technique* used a sheet-polarizing material, which was invented by the founder of the Polaroid Corporation, E.H. Land. In front of dual projection lenses and the viewing glasses, one vertically and one horizontally polarized sheet was mounted. The polarizers ensured that each eye saw its appropriate image of an image pair taken by a dual-camera system. This technique was commercially used for feature presentations in the 50s, when the film industry started competing with television. The down fall of the 3D movie industry was caused by different technical problems with the camera and projection system, camera crews lacking experience in stereoscopic photography, poor quality control in the film studios etc.

In modern times, stereo pairs of images are printed on a single roll of film in 35-mm over-and-under or 70-mm side-by-side format. Existing projector hardware can be used. Only the projector optic needs to be changed and a modulator is added to switch the characteristic of polarized light emerging from the lenses. The viewer wears passive glasses to view the projected image. This technique finds its application in some of the IMAX theaters around the globe.

Electronic displays offer the possibility to apply stereoscopy to applications other than found in the entertainment industry. Some of the problems of stereoscopic projection listed above do not exist when using electronic displays. After some unsuccessful attempts of implementing often impractical, technological schemes, J. Butterfield patented a system for industrial and medical applications in 1974. The system consisted of a CRT on which two images were displayed side by side and were viewed through a stereoscope. In the late 70s, electro-optical shutters were introduced to the market. They allowed left and right images to be shown in an alternating fashion when they were displayed on one monitor.

Many stereoscopic devices using either mechanical shutters, such as a spinning cylindrical mechanical shutter, manufactured by Bausch and Lomb, or electro-optical shutters, such as Kerr cell electro-optical shutters, which required very high voltage, were invented over the years. A major step forward to use stereoscopy as a productivity tool was the invention of the first modern electro-optical shutter, which was using lead lanthanum zirconate titanate (PLZT) ceramic, at Sandia National Laboratories in the early seventies. Later, at the beginning of the nineties, J. Roese experimented with these shutters at the Naval Oceans System Center in San Diego. He mounted PLZT shutters in goggles for a stereoscopic selection device. Although the stereoscopic images were dim and flickered, it was demonstrated that this technology had the desired capability to produce images with an impressive stereo depth. Megatek then applied his idea to computer graphics products. Developers in the company tried to overcome the flicker by interrupting the image by blinking the shutter during each field. This approach failed because parts of the image were chopped away in a horizontal direction and flicker was not reduced.

In 1981, StereoGraphics Corporation, headed by Lenny Lipton, started manufacturing a flickerless, field-sequential electro-stereoscopic display, which allowed to refresh a stereoscopic image 120 times per second, ensuring that each eye sees a flickerless image at 60 Hz. The originally used PLZT shutters, which were too dim, used too much voltage, and were too expensive, were replaced with bright, energy-efficient, liquid crystal (LC) lenses. The LC shutters were incorporated into a headband visor. But the visor was uncomfortable and had to be connected to a controller by a cable, which was difficult to accept by many users.

In 1984, StereoGraphics and Tektronix offered a solution to this problem by designing a front LC panel for the monitor, similar in appearance to an anti-glare screen. The panel changes the light emitted by the display tube of the monitor into polarized light whose characteristics change with each video field. Passive glasses with polarizing filters complete this system. Using circular polarization allowed horizontal head rotations without increasing the crosstalk between both projection channels. Such a passive system can be quite expensive for a system with a large monitor. This is the reason that StereoGraphics introduced its active CrystalEye system in 1989.

StereoGraphic's CrystalEyes system (Fig. 6) does not have an expensive LC panel but battery-powered glasses with achromatic LC lenses. The achromatic shutters guarantee fast transmission, low voltage use, fast transition time from closed to open shutter state, and a good dynamic range at low cost. An emitter sitting on top of the monitor sends out infrared signals to synchronize the LC lenses in the glasses and the images displayed on the monitor. It has more than four times of the dynamic range of the large LC panels, which improves image quality and reduces crosstalk between the channels. The electronics of this emitter convert the computer sync pulse into a code that modulates the emitted infrared radiation. A sensor, which is mounted between the two LC lenses in the eyewear, receives the signal, interprets it, and controls the opening and closing of the liquid crystal shutters, which are in sync with the display of the appropriate image. LC lenses are light, fast, work silently, and have no moving parts. The combination of eyewear and a high quality stereo-ready monitor with a refresh rate of 120 Hz or more is the most optimal hardware for stereoscopic applications available today.

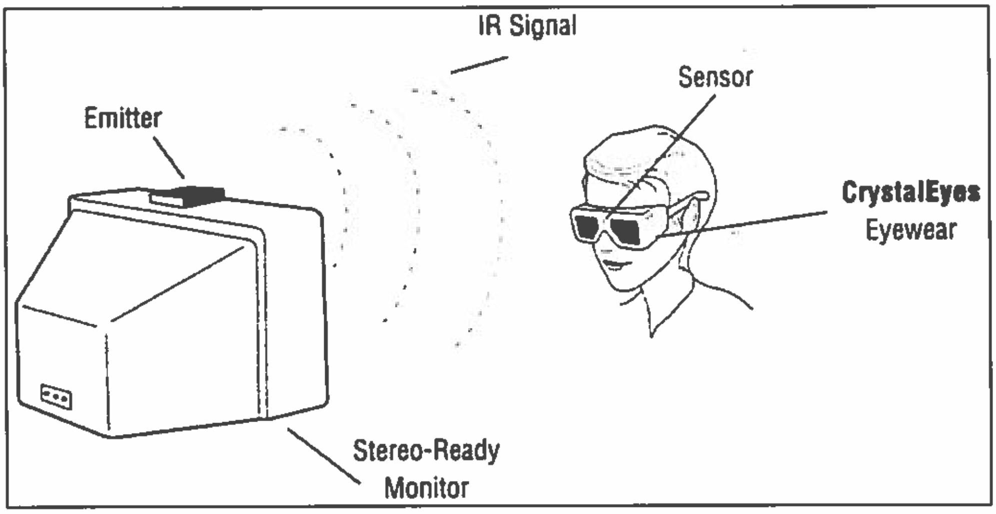


Figure 6: CrystalEyes setup
(source: Lipton, 1991a, p.336)

The immersive version of the eyewear, which we described above, is a head-mounted display (HMD). This approach involves two small flat-panel liquid-crystal displays (LCD) or two cathode-ray tubes (CRT), which ensure that each eye sees the appropriate image. LCDs are light, flat, inexpensive, and use little electrical power but had a very low resolution until Sony introduced a full-color LCD of 0.7 inches square that displays over 100,000 pixels. Developments in the CRT technology made it possible to manufacture 1-inch television screens that display clear, full-color images with a resolution of 1,000x1,000 pixels. Optics, mounted in front of the display devices, make sure that a comfortable fixed focal point is provided. However, when wearing an HMD, a viewer has no peripheral vision. The field of view for each eye does not exceed 70 degrees for high-performance HMDs. The goal is to develop the optical system in a way so that the user of these devices can experience a 180-degree field of view.

The fundamental principles of the design of stereoscopic viewing devices are applied to today's projection-based VR systems, such as the CAVE. The user of these systems experiences a virtual environment by utilizing display equipment, which is based on the principle of stereopsis that allowed Sir Charles Wheatstone to build his mirror stereoscope about 150 years ago.

3.2 Concepts for stereoscopic graphics

Most of today's computer-graphics generated displays do not exploit depth information to its fullest extent because they are created for conventional 2D screens. Much depth information gets lost when we project a 3D image onto a 2D screen. Furthermore, our skill to extract information with two eyes from a third dimension is not fully utilized and techniques, such as object translation and object rotation, are essential to explore this dimension in an originally three-dimensional environment. These conventional displays make use only of pictorial depth cues to provide the viewer with an image of a scene. In contrast, stereoscopic displays, or *true 3D* displays, can employ depth cues that are essential to perceive a true third dimension and, therefore, provide more detailed depth information that can be especially useful to view more complex objects and to understand the depth relationships between them. They often increase user performance for accuracy and reaction times for specific applications (see ch. 4). But it is known, when improperly produced and displayed, stereoscopic images can have an adverse effect. Poorly designed algorithms can introduce vertical parallax and spatial distortions (Hodges & McAllister, 1990). Below, we give an overview about hardware and software requirements and their interplay to create stereoscopic displays.

The stereoscopic hardware has to ensure that two disparate images, one for the left and one for the right eye, are delivered to each eye. The display creator can choose between two approaches: passive or active eyewear. Both techniques have in common that they use a stereoready computer monitor that is capable of a refresh rate of 120 frames per second so that two images, which are displayed sequentially, can be fused together and viewed without a disturbing flicker. Unfortunately, doubling the refresh rate reduces the resolution of each image display to the half. Furthermore, the apparent brightness of the display is reduced. This has two reasons. First, each eye sees the image as a sequence of image, nothing, image, etc. Second, each image has to pass through the lenses of the eyewear. Passive glasses have lenses that polarize 90° off from each other. In addition to the glasses, a polarizing LCD panel is used in front of the monitor screen, which polarizes images for the left and right eye in the same manner as the glasses do. Active glasses are infrared controlled LC shutter glasses that are synchronized with the computer display. Every time the left or the right image is shown on the display, the left or right lens of the glasses is transparent, meanwhile the other lens is opaque.

To generate stereoscopic images, no special rendering software is needed. Only one basic requirement has to be met by the software: it must produce a perspective view of a scene. After two slightly different views of a scene have been rendered, the views are loaded into display memory and are displayed one at a time. The software ensures that the images are toggled after the vertical retrace interval is detected by issuing a vertical retrace interrupt or by polling a line counter in the display controller. During this vertical blanking period, the lenses of the eyewear get synchronized as well.

There are two basic models for stereoscopic projection: *crossed axes* and *parallel axes* projection. The models can be explained using a camera metaphor. In the crossed axes method, the optical axes of the two cameras cross at a certain point in space. Using these camera positions, the images are then projected onto the image plane. A variation of this method, originally applied in computer graphics, is using rotation through some degrees to obtain one view from the other. As shown analytically by Hodges and McAllister (1990), this method introduces vertical parallax between the stereo image pairs. The rotation of perspective views results in a semicylindrical stereo window, i.e. the display screen. This curved stereo window distorts relative depth relationships in an image. As a result, a warped percept is produced. For example, displaying a rectangle in stereo should produce a rectangle appearing at some depth (Fig. 7a). Using rotation to generate the stereo image produces a curved percept (Fig. 7b). This percept is broken up into segments that appear to lie in different depth planes (Fig. 7c) because of the discrete pixel locations.

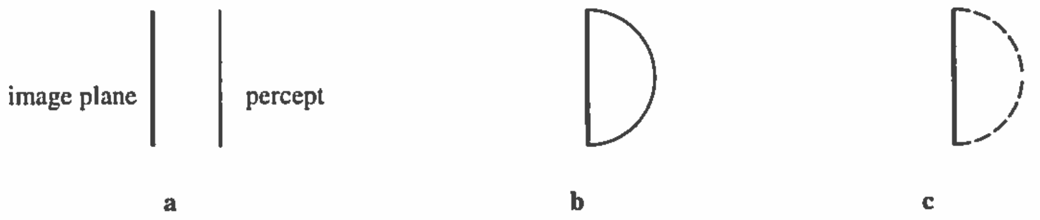


Figure 7: Distortions introduced by semicylindrical stereo window (adopted from Hodges and McAllister, 1990)

In the literature (e.g. Akka, 1991a, 1991b; Hodges, 1991), two basic approaches for the successful production of stereoscopic images are described: *parallel off-axis* and *on-axis* projection. As we show in the next paragraphs, both approaches are mathematically identical but generate a slightly different image pair. They can be implemented using standard viewing transformations such as described in Foley, van Dam, Feiner, and Hughes (1990, chapters 5 and 6) or Vince (1995, chapters 4 and 5).

The off-axis projection (Fig. 8) is based on two different centers of projection LCoP at (-e/2, 0, -d) and RCoP at (e/2, 0, -d), where *d* represents the distance between the coordinate system origin and the center of projection, and *e* represents the horizontal separation between the two centers of projection. One of the centers of projection is for the left-eye and one for the right-eye view. The centers of projection are horizontally aligned. Assuming a view plane that passes through the origin of a left-handed coordinate system and that is located parallel to the xy-plane, we can formulate the projected value of P(x, y, z), which can be any point in 3D space, for the left-eye view $P_L(x_L, y_L)$ and for the right-eye view $P_R(x_R, y_R)$

$$x_L = (xd - ze/2)/(d + z), \quad y_L = yd / (d + z), \text{ and}$$
$$x_R = (xd + ze/2)/(d + z), \quad y_R = yd / (d + z).$$

The fact that $y_L = y_R$ ensures us that no vertical parallax is introduced to our stereo image.

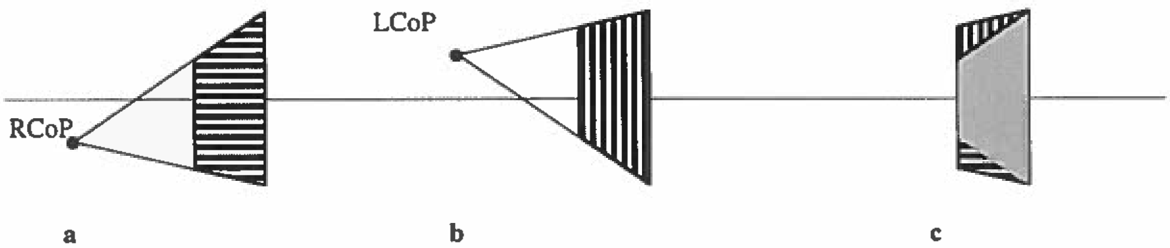


Figure 8: Off-Axis projection showing views for the right-eye, the left-eye, and both eyes

The field of view for the produced stereo image using the off-axis projection method consists of two monoscopic regions, which are regions only seen by the left or the right eye, and a stereoscopic region, which is a region seen by both eyes (Fig. 8c). The combined regions present a wider field of view than the one achieved using a standard perspective projection method for a single image.

The on-axis projection method is accomplished by horizontally moving the image plane by e/2 to the left or right and then producing the projected image (Fig. 9). Therefore, the left- or right-eye view is obtained by the following steps:

1. Translate the image plane by e/2 or -e/2 respectively
2. Project the image using standard perspective projection
3. Translate the image plane back by -e/2 or e/2 respectively.

Compared to the off-axis projection method, the on-axis projection method has advantages and disadvantages. The field of view of the on-axis projection still has two monoscopic regions and one stereoscopic region (Fig. 9c), as is the case for the off-axis projection. It is not larger than the one of a single perspective projection. Furthermore, the arrangement of the three regions is different compared to the off-axis projection. Translating the left- or right-eye image back to its original position results in a loss of data on one side of the screen and an empty field on the other side of the screen. The significant advantage of the on-axis projection is that it can be implemented using standard transformations and the standard perspective projection. At many graphics workstations, these operations are available in hardware and guarantee a better overall performance. In comparison, the operations for the off-axis projection must be implemented in software because only few workstations have graphics hardware that allows the programmer to control the position of the center of projection.

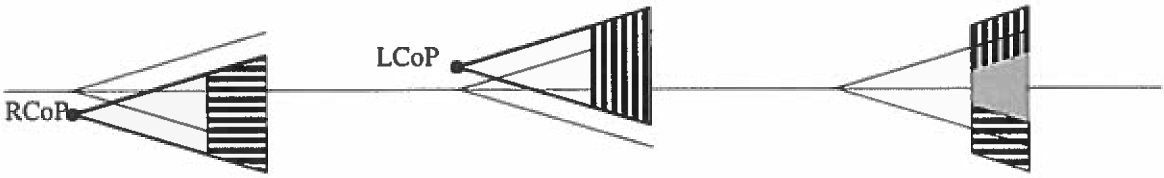


Figure 9: On-Axis projection showing views for the right-eye, the left-eye, and both eyes (adopted from Hodges, 1991)

The value of the parameter e, the horizontal distance between the two projection centers, is important because it determines if an observer of the stereoscopic image is able to fuse the left- and the right-eye view. According to Hodges (1991, p.12), the proper value of e depends on many circumstances, such as "the dominant color of an object, the location of an object on the screen, whether the object employs negative (crossed) or positive (uncrossed) horizontal parallax, the observer's own visual system and experience, distance of the observer from the screen, size of the display screen, and linear distance of an object from the projection plane". The amount of horizontal parallax depends not only on the value of e but also on the value of d as is shown in the following equation:

$$P = x_L - x_R = [d(x - e/2) \ / (d + z) + e/2] - [d(x + e/2) \ / (d + z) - e/2] = e\ (1 - d \ / (d + z)).$$

Because display screens vary from small to large projection screens, P is often expressed in terms of the visual angle

$P = 2d \tan(\beta/2)$ where β is the *horizontal visual angle.*

Choosing the appropriate value for β is critical for the successful generation of a stereoscopic image. If β is too small, the perceived depth in the image is small or lost. If β is too large the image is difficult to fuse. Valyus' guideline (Valyus, 1966) is *allowable parallax = 0.03 * viewing distance.* He reports an upper limit of 1.6° for the parallax angle to ensure comfortable viewing conditions. Akka (1991b) points out that this is not a firm figure. Lipton (1982) argues that "negative parallax values several times the rule given by Valyus are permissible for objects moving rapidly out of the frame of held on the screen briefly". But exceeding a certain limit causes uncomfortable viewing conditions and is called *accommodation-convergence conflict.* This parallax constraint limits, of course, the depth range of a stereo image and recommends a certain viewing distance. By adjusting the depth range of displayed objects and the viewing distance, the parallax angle can be kept within acceptable limits.

3.3 Stereoscopic displays—the first steppingstone to high-end VR systems

Bryson (1996) defines virtual reality (VR) this way: "VR is the use of computers and human-computer interfaces to create the effect of a 3D world containing interactive objects with a strong sense of three three-dimensional presence". According to this definition, VR is computer-generated, three-dimensional, and interactive. The goal is to present a high quality, graphical 3D scene to an observer to trigger a perceptual response similar to the response to a real scene. Thus, virtual reality tries to create an effect, and not an illusion of the real world. This effect can be obtained through the use of often head-tracked, usually stereoscopic, displays, high performance graphic systems, and an input device that allows users to interact with the system in three dimensions.

Given the definition above, VR systems can be divided into three groups: *non-immersive*, *hybrid*, and *immersive*, which we explain briefly in the next paragraphs. A *non-immersive* system, usually workstation-based, permits users to observe and/or navigate through a virtual environment of a 3D scene through the use of some special hardware (e.g. a electro-stereoscopic display), leaving users visually aware of the real world surrounding them. There has been a discussion, whether such stereoscopic system configurations should be called VR or not. Some authors (e.g. Vince, 1995) classify these systems as *through-the-window* VR systems. Others (e.g. Ware et al., 1993, p. 37) call systems, which are able to display "a stereo image of a 3D scene viewed on a monitor using a perspective projection coupled to the head position of the observer", *Fish Tank Virtual Reality*.

Hybrid VR systems, or *augmented reality* systems, allow the user to observe the real world with virtual images superimposed over this view. A practical example is found in head-mounted displays used by service engineers at Boeing, which allow the engineer to view their outside world, a plane undergoing service, simultaneously with overlaid synthetic graphics or text. Graphical and text information helps the engineer to locate parts and to service various components of the plane.

Using *immersive* systems, the user's view of the real world is replaced with a computer-generated virtual environment. The view of the virtual world reacts to changes in the position and orientation of the user's head. In such environments, the user is visually disconnected from the real world. The feeling of complete immersion can be further enhanced by the addition of sound and touch to interact with the virtual environment.

At SIGGRAPH 92, the CAVE(TM), the first Virtual Reality Theater that is a multi-person, room-sized, high-resolution, 3D video and audio environment, premiered. CAVE stands for Cave Automatic Virtual Environment. It and other high-end VR environments, such as an Immersa-DESK (VR drafting table) and an Immersa-WALL (large-screen projection display), were developed by the Electronic Visualization Laboratory (EVL) at the University of Illinois Chicago, which is a center for research and development of software, hardware, networking and communications tools for Virtual Reality (for more information see http://www.evl.uic.edu/EVL/VR/). EVL develops projection-based virtual reality hardware, software, and applications. Projection-based VR is similar as head-mounted VR with one big difference: the user doesn't carry the display equipment. Currently, four high-resolution projectors (1280x492) are used to produce full-color, computer-generated images for three walls (3 m x 2.75 m) and the floor. The software is able to support a 6-wall CAVE. It synchronizes all the devices and calculates the correct perspective for each wall. In the current configuration at EVL, one SGI Infinite Reality Engine is used to create images for the walls. Computer-controlled audio provides sound to a network of speakers. The sound is used as response to user actions, or when the application needs to signal the user about a change in the environment.

In the CAVE all perspectives are calculated from the point of view of the user. A head tracker provides information about the user's position. Offset images are calculated for each eye. The users wear Stereographics' CrystalEyes liquid crystal shutter glasses to see the virtual environment in three dimensions. The wand—a 3D mouse with a joystick and three buttons that can be programmed for interactivity—is used to navigate and interact with the virtual environment. The applied mathematics and depth cues allow the user to experience a three-dimensional virtual environment. The projection system allows more use of peripheral vision. Real and virtual objects are in the same space so that users have an unoccluded view of their own body as it interacts with virtual objects. One user out of a possible group of users is monitored by a tracking device to update the visual display with respect to the users' position and input. The viewer can request an enhanced non-real time rendering of a desired data subset. EVL's CAVE (VR theater) currently uses a Flock of Birds tracker to monitor the position of the user's head and wand-holding hand. The general interface was designed to be flexible enough to be used in a wide variety of applications.

The CAVE environment is used to explore visualizations of precomputed data sets. The user is able to explore interactively new scenarios, and experience a real-time visual response. A wide variety of applications, such as architectural walkthroughs, scientific visualizations, weather simulations, molecular and medical modeling, was developed to understand the kind of data that could benefit from implementation in a VR environment. Furthermore, user studies try to explore features that can be useful in developing VR applications. Cruz-Neira et al. (1993) came to some interesting conclusions in their report on visualization applications in the CAVE environment. They found that the inside-out paradigm, when viewing objects, was seldom applied. Most applications chose to view their data from a distance, like in a low-cost, non-immersive VR environment. Memory limitations of the CAVE seriously constrained the amount of data that could be visualized at one time. Bryson (1996) points out that large physical memory alone is not sufficient to solve all data management problems knowing that the size of data sets will increase dramatically over the next years. He suggests several strategies, such as subvolume loading, subsampling of the data, data compression, optimized data organization, and simplified graphics. Real-time visualization of large amounts of data in the CAVE asked for computing this data concurrently. Therefore, computationally expensive tasks had to be delegated to external computers and the data transmission was realized through high-speed networks. The study showed also that improvements to the hard- and software interface are critical. Most interactions used only simple movements and button presses of the 3D input controller, possibly because of problems in the design of the wand device.

3.4 How to interface virtual worlds

Interactivity is fundamental to scientific visualization because science as a method is by nature an interactive and iterative process. The section above focused on some aspects of the computer-related side of VR. The second component of these systems is related to the requirement of a natural, inherently three-dimensional interface, which also can incorporate sound and haptic input. Using conventional interfaces, which are based on text and two-dimensional input using graphical user interfaces and two-dimensional perspective displays of our 3D world, it is difficult to specify positions in three dimensions. Furthermore, these interfaces do not provide an unambiguous display of three-dimensional structures. Therefore, VR attempts to provide a more anthropomorphic 3D interface. Well-designed three-dimensional interfaces and stereoscopic displays promise to enhance significantly the ability to explore rapidly and easily a complex data environment. In this section, we, first, present a brief and high-level discussion of 3D interaction devices, interaction tasks, and interaction techniques. We, then, explore how we can interact with virtual worlds.

3.4.1 Three-dimensional interaction devices and interaction tasks—the basics

3.4.1.1 Three-dimensional interaction hardware

Interaction devices are pieces of hardware by which the user enters information into a computer system. A tremendous variety of such devices with an abundance of designs exist on the market today. Therefore, we concentrate in our discussion on selected input devices specifically oriented towards 3D interaction. Many strategies were developed to organize the different designs. Foley et al. (1990) used basic logical device categories, such as *locator*, *keyboard*, *choice*, *valuator*, and *pick*, in their discussion of input devices. Buxton (1983) proposed a classification scheme based on the physical properties and the number of spatial dimensions of input devices. But he limits the discussion to continuous devices. Card, Mackinley, and Robertson (1990, 1991) systematize knowledge about input devices by defining the design space based on physical properties sensed by input devices (Table 4). They extend Buxton's taxonomy by introducing the notion of a composite device and by representing discrete devices. Composite devices are formed by taking the cross product of the input domains of two devices to define the input domain set of the combined device (merge composition), by collocating two different devices in different places of a common panel (layout composition), or by mapping the output device of one device onto the input domain of another (connect composition). This taxonomy covers the common input devices with the exception of voice and heat sensors.

<table>
<tr><td></td><td colspan="3">Linear</td><td colspan="3">Rotary</td></tr>
<tr><td>Absolute position</td><td colspan="3">Position P</td><td colspan="3">Rotation R</td></tr>
<tr><td>Relative position</td><td colspan="3">DeltaPosition ΔP</td><td colspan="3">DeltaRotation ΔR</td></tr>
<tr><td>Absolute force</td><td colspan="3">Force F</td><td colspan="3">Torque T</td></tr>
<tr><td>Relative force</td><td colspan="3">DeltaForce ΔF</td><td colspan="3">DeltaTorque ΔT</td></tr>
<tr><td></td><td>1</td><td>100</td><td>Inf.</td><td>1</td><td>100</td><td>Inf.</td></tr>
<tr><td></td><td colspan="6">Measures</td></tr>
</table>

Table 4: Physical properties recognized by input devices
and the measure of the domain set, i.e. number of values sensed
(adopted from Card et al., 1990, p. 119 & 120)

Let us now discuss some common 3D-interaction devices. Some of them are 2D devices extended to 3D, such as mice, trackballs, and joysticks. We can think of the 3D trackball as a mechanical mouse lying on its top. Potentiometers or shaft encoders sense the motion of a freely rotating ball inside of a housing. By rotating the ball with their palms, users input rotational information. The trackball is made to sense rotation about the horizontal and vertical axes. The 3D joystick has a stick that can be moved left or right, forward or backward, or can be twisted clock- or counterclockwise. Again potentiometers sense the movement. Logitech's 3D-mouse system allows input of three-dimensional, spatial information. It consists of three major components: a 3D-mouse transmitter, a 3D-mouse receiver, and a control unit. The 3D mouse transmitter sends 23 kHz ultrasonic signals to the receiver that relays these signals with regard to its position and orientation to the control unit. The control unit decodes the signals from the 3D-mouse receiver and computes the receiver's position and orientation. This unit then reports the data and mouse button activity to the host computer.

Figure 10: *The Magellan/Space Mouse (left) and schematic presentation of working principle (right)*

The Magellan/Space Mouse (Fig. 10, left) and the Spaceballs 2003/3003 allow the control of 3D graphical objects and images in six degrees of freedom instantaneously and simultaneously. Graphical objects are moved due to a translational or rotational displacement of the Space mouse cap or the Spaceball's Powersensor ball. The cap and the ball are suspended by leaf springs. Cap or ball movements will cause a deflection of each leaf spring (Fig. 10, right), which is measured by opto-electrical measurement system inside of the devices. A detector, which is part of the measurement system, produces a current that is in proportion to the amount of deflection. This current is then converted into a digital value. The data glove is able to record hand position, orientation, and finger movements. The glove is covered with fiberoptic cables to sense finger movements. An absolute position and orientation sensor helps to record hand movements. The fingertips are often covered with a tactile-feedback device. A user wearing the data glove can grasp objects to move and rotate them. We are very familiar with these interactions from our real world experience. In Table 5, we plot these input devices on the taxonomy by Card et al.

	Device name	Discrete components	Linear			Rotary		
			X	Y	Z	RX	RY	RZ
Absolute position	VPL data glove	(10)	●	●	●	●	●	●
	Logitech's 3D mouse	(5)	●	●	●			
Relative position	3D joystick					●	●	●
	3D trackball,					●	●	●
	3D Spaceball 2003/3003,	(11)	●	●	●	●	●	●
	Magellan/ Space Mouse	(9)	●	●	●	●	●	●
			1...n	1...n	1...n	1...n	1...n	1...n
			Measures					

Table 5: 3D input devices plotted on a taxonomy based on the taxonomy by Card et al. Black lines indicate merge composition. Dashed lines represent a layout composition.

3.4.1.2 Three-dimensional interaction techniques

Let us define an *interaction task* as an entry of information by the user (Foley et al., 1990). We call the implementation of an interaction task, as a sequence of input device actions, *interaction technique*. Hence, there exist many different interaction techniques for a given interaction task. Two interaction tasks, *position* and *select*, are tasks we find in 2D applications as well. A third task, *rotate* or *orientate*, is specific to a 3D environment. One problem of interacting in 3D space arises because 2D devices are often used and their movements in 2D must be mapped into 3D. We will illustrate this problem discussing positioning and rotation techniques in the next two paragraphs. Newer devices (see 3.4.1.1) try to overcome this problem by providing a more natural and intuitive method for moving in 3D space. Hence, these devices allow the user to concentrate on the interaction in 3D and not on the technique that is required to accomplish this interaction. Another difficulty arises from the fact that the user needs to perceive the 3D depth relationships of objects in the display correctly in order to interact with these objects. We consider stereoscopic displays as a potential way of improving the understanding of depth relationships in 3D displays.

Foley et al. (1990) suggest several ways to map 2D device movements into 3D movements. Most of these techniques decompose the 3D task into simpler tasks of a lower dimension. One technique uses three different views of a 3D scene, for example front view, side view, and top view. The user can specify any 3D point in the scene by dragging one or two 3D cursor lines, whose intersections in each view specify a 3D cursor location. On one hand, multiple views can be advantageous because they simplify object relationships in 3D space. On the other hand, this technique restricts the user to work at one or two views at a time. Another way to implement 3D positioning is to allow cursor movements only in the directions of the projections of the three principal axes in a coordinate system (for example Balakrishnan et al., 1997). This technique, of course, restricts movements to one axis at a time. Sometimes, the user will be allowed to place a local coordinate system on the surface of an object in a 3D scene. This coordinate system defines the possible directions of movement for the selected object.

In the case of 3D rotation, we again decompose the 3D task into 2D tasks. First, we establish the direction of the three axes. This should be done in accordance with the screen coordinate system. Second, the center of rotation must be explicitly or implicitly specified, for example the origin of the screen coordinate system. Third, dials or sliders can be used to control the rotation about the x, y, and z axes. Another way of implementing the third step is to map mouse movements directly onto scene or object movements. Horizontal and vertical mouse movements could be mapped to control rotation about the x and y axes. Chen, Mountford, and Sellen (1988) extended this technique. They implemented z-axis rotations by pressing one of the mouse buttons and moving the mouse. As a metaphor, they ask the user to imagine a superimposed 3D trackball on top of the object. In another method also developed by Chen et al., circular mouse movements control the rotation about the z axis.

Three-dimensional devices try to overcome some of the problems mentioned in the paragraphs above. The user easily manipulates objects in a 3D scene by pushing, pulling, or twisting the cap of the Magellan 3D controller or the ball of the Spaceball 2003/3003 in any direction. Pushing or pulling of the device will cause a scene translation. Twisting will rotate a scene. The Magellan as well as the Spaceball is used in conjunction with a mouse. Hence, the user positions an object with the 3D device and operates on the object using a mouse. These 3D devices, including Logitech's 3D mouse, provide directly X, Y, and Z values specifying the amount of translation and rX, rY, rZ values specifying the amount of rotation. Although maybe still imperfect with regard to their design, we believe that 3D devices simplify interactive 3D control by providing a more natural interactive control over the 3D scene. Using 3D input devices, Hinckley, Tullio, Pausch, Proffitt, and Kassel (1997) showed that users completed a 3D matching task up to 36% faster without sacrificing precision.

3.4.2 Interacting in 3D space

Keeping in mind the above discussion of 3D interaction hardware and techniques, we now turn to some of the problems of interacting in virtual environments. One vital component of a stereoscopic visualization system is the display. High-quality, high-resolution, full color displays with a 1000x1000 pixel resolution are sufficient for many scientific visualization applications (Bryson, 1996). In these applications, a wide field of view is often desired to explore not only the fine details of a structure but also the more global phenomena. The combination of high resolution and wide field of view is extremely difficult. Alternatives, such as projection-based systems, seem to be promising for scientific visualizations. As Brooks (1986) points out, the full power of immersion VR is only needed for a few applications, such as architectural walkthroughs or entertainment. Hence, fish-tank VR systems have been proven to be sufficient and more cost effective. These systems are able to render the three-dimensional scene from the actual position of the user's head. Once the head position is known, the view can be rendered in stereo from the user's point of view. Computer graphics can provide additional depth cues, such as occlusion or lighting and shading. It is important to emphasize that a careful domain analysis is required to decide for one or the other VR system. Therefore, we need guidelines that support a potential user in the decision process. In addition, the user acceptance is an important issue when selecting the appropriate visual display. Many people express their distaste for wearing helmets, special glasses, or attaching displays onto their heads.

If the virtual environment should ease the process of data exploration and discovery in scientific visualization then there is clearly a need to control the visualization display interactively and intuitively, that is to quickly and precisely select a arbitrary location in 3D space. Bryson (1996) specifies three components to this ability:

- specify a location,
- specify an action at that location,

- provide feedback as to the selected location.

One way, and maybe the most intuitive one, to specify a location in three dimensions, is to track the user's head position through available tracking technologies. Unfortunately, these technologies suffer from a variety of problems, such as cost and lack of accuracy. Currently, there is no standard way of specifying a location in three dimensions. It is hard to imagine that the use of a two-dimensional mouse or extensions of it (Venola, 1993; Balakrishnan et al., 1997) is more than an intermediate solution to the problem. It is not as fast and easy as the use of a three-dimensional tracker. Furthermore, we question if a two-dimensional pointing device has its place in a three-dimensional world.

Of course, conventional hand-held button devices can also be used to specify an action at the chosen location. Although button devices are unambiguous and easy to learn, they must be held in the hand and provide an arbitrary method of specifying interaction commands. Fatigue problems are likely with the long-term use of such devices in virtual environments. Glove devices are an often used alternative to this approach. They can measure the angle of bend of the user's finger to determine a particular gesture out of a fixed vocabulary, such as 'fist', or pointing with one or more fingers. Although gloves provide an intuitive control of the virtual world, it seems doubtful to us that they always will be a part of future VR systems in their current form and use because the provided measurements are often inaccurate. Furthermore, there is a lack of a standard vocabulary of intuitive gestures. When data gloves are used in current systems, there is a need to calibrate out the inaccuracies of measurements. It is suggested to use the glove with low arm positions to avoid fatigue and to provide appropriate visual feedback for the gesture recognition.

Mercurio and Erickson (1990) pointed out several interface problems in virtual environments, for example the use of non-intuitive gestures and inappropriate metaphors. They emphasized that none of the problems was inherent in the interaction techniques. In some situations, gesture controlled interfaces can pose new, unexpected problems, when they are not carefully chosen. For example, using the metaphor of flying through a 3D model, participants of an experiment flew sometimes inadvertently because of putting a finger on the chin or stroking the mustache because the gestures used in the experiment were natural ones. These gestures had no meaningful connection with the commands they invoked. Furthermore, flying seems to be a poor metaphor system for movement through a virtual reality environment. Pointing at an object does not make users feel that they move toward the object; rather they feel as if the object is coming towards them because the kinesthetic feedback generally associated with movement is missing. We, too, believe that the purpose of an interface metaphor is to allow users to apply some of their real-world experience to areas of the new domain with which they might have difficulties otherwise. But often users do not know how to fly in the real world. Hence, they do not know intuitive gestures for flying. We think that the research of Mercurio and Erickson (1990) emphasizes the importance of an appropriate control metaphor because the types of manipulation, which the control metaphor affords, often determine how intuitively and how well the user performs in 3D environments.

Although some tasks in a three-dimensional environment seem to suggest an obvious way to interact with the visualization, this is not true for other tasks, which lack a well-designed interface. It is easy to change the location of a visualization by grabbing and moving it. But it is more difficult to come up with an idea how to select objects or how to change their state. One temporary solution to this problem might be the extension of conventional graphical user interface techniques, such as buttons, sliders, or menus. But what is really needed seems to be a new interface metaphor for the virtual 3D worlds such as the window metaphor was in the 2D world. In addition to the *flying metaphor* and its derivatives, e.g. the *locomotion metaphor* or the *car driving metaphor* (Brooks, 1986), many other metaphors for 3D interaction have been suggested: *camera metaphor* (Brooks, 1986), *object-in-hand metaphor* (Ware and Osborne), *ray casting metaphor* (Hinckley et al., 1994). Unfortunately, often the metaphors prove useful only under specific circumstances. For example, Ware and Osborne suggest that the object-in-hand metaphor is appropriate for manipulating closed objects, but not for walking through the object interior. It would be interesting to further investigate how task-specific 3D computer interfaces, in general, and interaction metaphors, in particular, are.

An obvious extension to the interaction modes discussed above is the inclusion of haptic feedback, that is feedback using the senses of touch and force, in the virtual environment. This allows the user to touch (and feel) the virtual object. As of today, it is unrealistic to expect 3D visualization to provide a similar tactile feedback as experienced by us in the real world. But there are tactile gloves available that activate small pads along the fingers to simulate the sensation of touch of a weightless object that is displayed in 3D space. Force feedback is used on the flight controls in a flight simulator allowing the pilot to encounter forces that are experienced in real planes. Similar manipulators can be used to create forces, such as physical mass of an object in the virtual world.

Three-dimensional sound becomes a natural feature to complement visual and tactile feedback from the 3D visualization. Although relatively unexplored, it holds promise (Minghim & Forrest, 1995). The conventional use of sound is to provide user feedback about the state of the environment, for example to signal object collision or to accentuate the interaction. State-of-the-art hardware is already attempting to simulate the way our ears perceive sound in the real world by supplying signals that model the attenuation of pressure wave entering the user's ear channel. We think, it is important to explore in more detail the use of sound as a data display channel. The proper integration of sound in visualization tools could ease the workload of our already overloaded visual system. Mapping scalar quantities to the frequency, amplitude, and timbre of sound can accomplish this.

In this chapter, we have discussed the composition of stereoscopic displays including some aspects of interacting with them. We have shown that the human physiology and psychology play an important part in this procedure. Having emphasized that the user is a central part of the stereoscopic visualization process, we now focus on some aspects of human factors research in virtual worlds by discussing effectiveness and user acceptance of stereoscopic displays in the next chapter. We address also perceptual artifacts based on limitations of the display hardware and the human visual system.

4 Visual Perception—the Human Factors of Visualizations in 3D

VR applications, such as stereoscopic visualizations, ask for, in many aspects, a completely new interface for human-computer interaction that is partly caused by the introduction of new hardware. Conventional input devices, such as the keyboard and the 2D mouse are replaced by gloves, 3D pointing devices, and voice input, and conventional monitors are substituted by stereo-ready monitors, projection-based stereo systems, or head-mounted displays. Therefore, the design of interfaces for virtual environments asks for the replacement of old affordances and metaphors (e.g. windows) by more appropriate ones to enhance task performance in virtual worlds. Hence, systematic research is needed addressing human factors issues to make VR systems more effective and welcomed by the user.

4.1 Human depth perception and performance in 3D environments

In the next three sections, we summarize several factors that affect depth perception in stereoscopic displays such as the type of disparity, the presence or absence of a number of monocular and binocular depth cues, and the wavelength of color. Taking a closer look at studies that examine depth perception and performance in virtual environments, we want to give some incentives why the use of stereoscopy in different applications can be advantageous.

4.1.1 Stereoscopic depth and the range of disparity

One fundamental concern regarding stereoscopic depth perception is how correct is the depth we perceive. We discussed in 2.1.1.1 that the range of disparity is divided into two directions relative to the plane of fixation. Research shows that there exist two separate mechanisms for processing crossed and uncrossed disparities and that about 30% of adult observers are insensitive to disparity of only one direction (stereoanomaly).

Patterson et al. (1989, 1992) studied the stereoscopic depth sensitivity in random dot stereograms using crossed and uncrossed disparities. This study shows that depth is more veridical and the duration at which depth becomes apparent is shorter with crossed than with uncrossed disparity. A threshold was defined as the shortest duration at which depth was perceived. These thresholds were lower, which indicates a greater sensitivity, with crossed than with uncrossed disparities. The amount of depth perceived at thresholds was very different across observers. The perceived depth was compared to a predicted depth, which is given by

$$d = S \times D / (I + S) \text{ for crossed disparity and } d = S \times D / (I - S) \text{ for uncrossed disparity,}$$

where d is the predicted depth interval between object and fixation plane, S is the separation between monocular half-images of the disparate stimulus, D is the viewing distance from observer to fixation, and I is the interpupillary distance. For crossed disparity, the range of depth estimates was close to the predicted depth value. For uncrossed disparity, the range of estimates felt far short of the predicted value. Furthermore, the perceived depth with crossed disparity is much better to that with uncrossed disparity indicated by lower thresholds and fewer depth reversals.

The studies show that this asymmetry reflects the sensitivity of mechanisms that compute perceived depth from disparity information (Patterson et al., 1989, 1992). It is suggested that differences exist between crossed and uncrossed mechanisms at levels of the visual system where perceived depth is computed but not at the level where disparity is detected or recognized. When exposing the stimulus for unlimited time, perceived depth was slightly greater than depth at a threshold for both directions. The results of these studies are very important for the design of stereoscopic displays. They suggest using crossed disparity, or a negative value of parallax to convey depth. When using uncrossed disparity or a positive value of parallax in the display, additional depth cues should be added to the display.

Lipton (1991b) points out a cue conflict arising when crossed disparities are favored in the display. As we will discuss further in 4.3, interposition cue and crossed disparity might compete with each other. In their recent research, Becker and Patterson (1997) address the issue of competing depth cues. They suggest that the difference between crossed and uncrossed stereopsis may be due, at least in part, to the depth relationship of the targets to the occluding surface of the background dots. The depth discrimination of uncrossed targets was only diminished when the stimuli appeared behind the background dots of the stereogram, that is the stimulus appeared window-like in the plane of background dots so that its boundaries could be interpreted as part of the background and not as part of the stimulus. However, in the case of crossed stimuli, a well-defined region appeared in front of the background.

4.1.2 Presence or absence of depth cues and user performance

Virtual worlds offer a greater flexibility than more traditional human-computer interfaces in task domains such as medicine, telecommunication, education, engineering, and, of course, entertainment. Scientific areas such as visualization, modeling, and astronomy benefit from virtual environments as well. It was a simple stereo-camera system that helped us to spatially explore one of our neighbors in space and discover a new world more efficiently. VR technology allows us to represent applications in three dimensions and to interact with them, to make use of egocentric and exocentric (i.e. birdeye's 3D viewpoints), and to choose stereoscopic or monoscopic views. Furthermore, it makes multi-sensory information in dynamically changing or relatively static displays available to us. But we also should be aware of current limitations of VR systems that restrict their usefulness and degrade the user's performance (see 4.2). Depending on the system and the application, these limitations include poor spatial-temporal resolution of visual, auditory, and haptic displays, interocular crosstalk, as well as inaccuracies and lag of head and eye tracking devices.

Human spatial perception is very important in all applications of virtual worlds. Ambiguous information in monoscopic displays often can be disambiguated by using multiple visual cues as provided for example in stereoscopic displays. Other sensory modalities can also be used to resolve this ambiguity. It is important to realize that the best perceptual cues and modalities are often task dependent. Therefore, we have to question for a specific application which perceptual information is critical, how this information should be conveyed to the user, and what simplifications are acceptable. In the next paragraphs, we summarize the results of some specific studies.

Although relatively little research has been done to determine when stereoscopic displays are more advantageous than perspective displays for spatial tasks, it seems that binocular disparity is particularly useful as a depth cue when monoscopic depth cues are either absent or degraded in some way in the image. Gallimore and Brown (Gallimore & Brown, 1993; Brown & Gallimore, 1995) studied the effect of stereoscopic depth information on the ability of CAD users to visualize 3D computer-generated objects. They wanted to determine the affect of presence or absence of an increasing number of monocular, such as interposition, texture, shading, and binocular depth cues. They found out that the influence, or the weight, of cues may depend on a specific task and the participant's strategy to solve a task. It was shown, that the addition of stereopsis does not necessarily lead to an improved performance in a specific task, but it can augment monocular cues sometimes, or it can disambiguate depth information. These results emphasize the importance of a careful analysis of depth cues with regard to their dominance in a certain display. Furthermore, the researchers concluded from their studies, that, although stereo images can be fused in 200 msec, it might take longer until we see them floating in depth so that the user can perform useful operations with the 3D object. It was highlighted that the disparity cue might be useful and even necessary during real-time interactive design because it not only enhances the visualization of complex 3D objects but also simplifies other design activities, such as the movement of input devices in the third dimension.

Similarly, Sollenberger and Milgram (1991), who compared rotational and stereoscopic displays, found that one depth cue alone (e.g. rotational motion or binocular disparity) was effective enough. Although a second, additional depth cue improved performance, rotational motion was a more powerful depth cue than binocular disparity, confirming Braunstein (1986), who showed that motion enhances effects of binocular disparity. But rotational motion was not better than the combination of binocular disparity and multiple viewing angles. This leads to the conclusion that instead of using expensive rotational display systems, it is appropriate to use cheaper stereoscopic display systems enhanced with software that allows graphic images to be displayed from multiple discrete viewing angles.

Arthur, Booth, and Ware (1993) looked at different combinations of head-coupled and stereo displays. They measured the user's performance on a 3D tree-tracing task, which was similar to Sollenberger's and Milgram's experiment. Participants of the experiment had to decide whether a leaf was part of a left or a right tree. They found that the head-coupled stereo condition was the fastest, but head coupling alone was slow. It was concluded that head coupling and stereo contribute to performance. The contribution of head-coupling to an increase in performance was larger than the one of stereo, which is probably due to a motion-induced depth. A closer look at the decision errors, which ranged from 21.8% in the static nonstereo condition without head coupling to 1.3% for the head coupled stereo condition demonstrates impressively the different influence of depth cues on user performance.

In general, a criterion for effectiveness of a certain display is determined by the immediate performance of a user as described above. But in visualization, we are also interested in long-term comprehension. When comparing monoscopic and stereoscopic displays with a high and low object-density, we would ideally like to show that stereoscopic displays reduce the visual scanning and search time within the more integrated rendering. Research conducted by Wickens, Mervin, and Lin (1994) studied cognitive processes, such as search and comparison, to clarify if techniques supporting short-term comprehension also support long-term comprehension. They showed that the integration level of the display influenced speed and accuracy of the performance significantly. Furthermore, they demonstrated that stereo has a significant benefit, which is enhanced at more complex and higher integration levels. Their data show that the benefits of stereopsis were directed more towards a global understanding of relations between points rather than local judgment of locations of specific points.

In a different task domain, Zenyuh, Reising, Walchli, and Biers (1988) compared how spatial location information about friendly, enemy, and unknown aircraft in a given volume space is provided by a conventional 2D display and a stereoscopic display. Results showed a significant accuracy performance advantage for those formats using the stereoscopic 3D display. In detail, accuracy was approximately 20% greater using stereo 3D displays and was approximately 20% better using size cuing, but accuracy was lower at high-density level displays. The difference in accuracy between 2D and stereo 3D displays increased with increased display density. The objective accuracy data support the prediction that retinal disparity and familiar object size yield more accurate search task performance in stereo 3D displays compared to the performance in 2D displays.

Muehlbach, Boecker, and Prussog (1995) looked into the future with their research. They compared monoscopic and stereoscopic representations regarding the immediacy of the spatial impression and perception of proportion, distance, and size of people and objects in videoconferencing. They expected that stereopsis would enhance the separation of a scene into foreground and background, and that, based on this fact, gestures and posture of participants of the videoconference may be better recognized. The visual impression of the interlocutor, who was at a remote site, was judged as looking more spatial and having more depth in the stereoscopic representation. Participants of the experiment stated in questionnaires that stereoscopic representations of the conferees enhanced the spatial presence as well as the telepresence, that is the degree to which participants of a videoconference get the impression of sharing space with interlocutors at a remote site. Furthermore, they affirmed that the use of stereoscopic displays made videoconferencing more attractive in terms of enjoyment and fascination. Stereopsis improved the estimation of size when compared with a monoscopic viewing situation, where interlocutors seemed to be smaller. It is interesting that 90% of the participants felt disturbed by the stereo glasses. This seems to be the main reason that there were no significant advantages of the stereoscopic system over the monoscopic system with regard to user acceptance and overall satisfaction. But this problem seems to be more or less caused by the current stereoscopic display technology and adds another point to the limitations of this technology as discussed in section 4.2.

4.1.3 Stereoscopic depth and the wavelength of color

In this section, we return to a depth illusion called *chromostereopsis* (see 2.2). McClain (1989) examined the effects that hues have on the perception of depth using stereoscopic displays. He was interested in monitoring the level of accuracy of participants with which they determine the relative depth of differently colored, adjacent symbols using varying combinations of hue and disparity in a quick-glance situation (time: 500 msec). While one symbol having the neutral hue of green and zero disparity acted as control symbol, the hue and disparity of the other symbol was changed. The experiment demonstrated that hue, disparity, and their interactions affected significantly stereoscopic depth perception. At the lowest disparity level (1 pixel separation), participants were less accurate in estimating depth differences, that is participants had greater difficulty to discriminate smaller disparity differences. At the maximum disparity level (3 pixel separation), the effect of hue began to decrease. The blue effect of the chromostereoscopic theory was confirmed in the experiment, that is with positive disparities the blue hue enhanced the sense of depth and accuracy improved; with negative disparities, blue decreased the sense of depth. In contrast, red supported the negative disparities and counteracted positive disparities. Green can be considered a neutral color because it is near the center of the hue frequency spectrum.

The described research has several implications for the design of stereoscopic displays. It suggests exercising caution when selecting hues, especially hues near the end of the frequency spectrum. When working with small disparities, we should avoid the use of colors at the extreme end of the frequency spectrum unless we want to enhance or weaken the intended depth effect in the display. For greater disparities, the effect of chromostereopsis seems to vanish. In addition, Owens and Leibowitz (1975) reported that the size of the illusory depth effect varies with slight changes of the interpupillary-distance setting of binocular instruments, such as microscopes. This represents a problem for stereoscopic software design as well. As in the case of optical-instrument design, small variations of the interpupillary distance can result in changes of the amount of perceived depth. Therefore, good stereoscopic display software has to allow different settings of the interpupillary distance.

Finally, we want to mention the interaction of choosing a hue for a field-sequential stereoscopic display and interocular crosstalk between the images of a stereo pair in a real system. Ideally, in field-sequential displays, we want that when one of the images is on the screen the other is extinguished and that this image is displayed only to one of the eyes. One factor influencing interocular crosstalk is phosphor persistence and, connected with it, the position of a figure in the stereoscopic image. Phosphor persistence is defined as the amount of time it takes for the phosphorescence to decrease to 1/10 of its initial output. It is known that green phosphor decays much slower than red and blue phosphors. A long phosphor decay rate is especially noticeable on the bottom of the screen because the scan lines there have less time to decay then scan lines at the top of the screen when the shutter switches. Therefore, interocular crosstalk progressively increases from the top of the screen to the bottom of the screen. In the next section, we will continue our discussion of problems related to depth perception in virtual environments.

4.2 Problems for depth perception in stereoscopic environments

It is important to realize that there is a difference in how we look at objects in the real world and how we look at stereoscopic displays regarding the accommodation/convergence relationship. In the real world, accommodation and convergence correspond, that is our eyes converge on an object and they also focus (accommodate) on it. When we look at stereoscopic displays, our eyes will converge at objects at different distances based on screen parallax, but they will remain focused on the surface of the screen. This deviation from the normal habitual response could cause some viewers of stereoscopic images to feel uncomfortable. This is especially true for stereoscopic images with large horizontal parallax values. Therefore, Lipton (1991b) suggests to generate images with the largest depth effect but with the lowest parallax values, and to keep the image near the plane of the screen. Furthermore, depth in the image can be enhanced by emphasizing the perspective cue, for example by increasing the angle of view, or by depth cueing, that is the reduction of intensity along the z-axis. This reduces a possible accommodation/convergence breakdown, which can be especially evident for small screens viewed from a close distance. The breakdown is less severe, when looking at large-screen displays from greater distances, but it cannot be overcome with the current technology.

Systems that respond to where the observer is looking in the virtual environment could be a technical solution to this problem. Wann (1995) proposes a system that is able to monitor eye movements and uses this data to adjust the image plane depth via a servo-lens system. Such a system requires a high-resolution display and real-time behavior of the eye-monitoring system and servo control without limiting the field of view or head movements. Presently, no current system is able to meet these requirements.

Lipton (1991) suggested setting the plane of zero parallax at the center of an object. When changing the scale of an image of an object, the zero parallax plane should be kept at the object's center. When displaying a more complex scene, the zero parallax plane should be placed at the nearest foreground object. This approach avoids negative parallax. The positive parallax has to be carefully controlled for distant objects by changing the distance between the two centers of projection and/or the angle of view.

When displaying images with negative parallax, it is best to avoid situations where the screen edge touches or cuts displayed objects because there will be a conflict of two depth cues: interposition and binocular disparity. One depth cue, the interposition cue, suggests that the displayed object is behind the display screen when it is cut off by the screen edge. The other cue, horizontal parallax in the image, suggests that the object is floating in front of the window. This situation of cue conflicts is unacceptable for most of the viewers. Lipton (1991b) points out that especially the vertical edges are troublesome. When looking through a window and comparing our left- and right-eye view we realize that, for example, our left-eye image shows more on the right window edge than our right-eye image. This experience cannot be replicated when we look through our stereo window, the display screen. In a stereoscopic image with negative parallax, our left-eye view, when looking at the right vertical surround, shows less of the image than the right-eye view.

The rule of thumb is that parallax values shouldn't exceed angles of much more than 1.5°. This corresponds to positive or negative parallax values of 12 mm when viewed from a distance of 432 mm. Akka (1991b) and Lipton (1991b) state that this rule was made to be broken and that it is best to let our eyes be our guide. Most important is to keep the distance of the observer of stereoscopic displays in mind when generating stereoscopic images.

Another factor influencing interocular crosstalk in field sequential displays is the dynamic range of the LC shutter glasses. The dynamic range quantifies the amount of leakage of the shutter in its closed state and is defined as ratio of the transmission of the shutter in its open state to the transmission of the shutter in its closed state. It is obvious that the amount of light leaking through a closed shutter is directly related to the intensity of unwanted ghost images. Hence, ghosting is produced by crosstalk between the left and the right image. Lipton (1991b) pointed out several additional factors that affect the perceived ghosting such as image brightness, contrast, textural complexity, and horizontal parallax. The following rules of thumb should be used to improve image quality:

1. Ghosting is directly proportional to brightness, contrast, and the amount of horizontal parallax.
2. Ghosting is inversely proportional to textural complexity in the image.

Testing for binocular fusion limits and evaluating the accuracy of depth discrimination in stereoscopic displays, Yeh and Silverstein (1990) showed that interocular crosstalk inconsistently effects diplopia thresholds, depending on the stimulus exposure. They also showed that interocular crosstalk basically did not affect the depth discrimination performance, but significantly influenced participant's ratings of image quality and visual comfort. The quality of red test images, causing low crosstalk, were rated higher than the quality of white images, which caused a high amount of crosstalk because the white color contains a large green component. Their data suggested that interocular crosstalk influences the development of eyestrain and headaches because the subjective rating of visual comfort was poorer for images displayed in white instead of red.

The CRT refresh rate determines whether a stereoscopic image is flicker free. Shutters and monitors working with a 60 Hz refresh rate are only able to display the left or right image at a 30 Hz refresh rate. The images are usually drawn into a hardware double buffer at full resolution. Increasing the refresh rate to 120 Hz allows us to refresh one of the images at 60 Hz. This approach though reduces the vertical resolution by half because each frame buffer is divided into a frame buffer for the left and right eye view. Unfortunately, as shown by Utsumi, Milgram, Takemura, and Kishino (1994), low-resolution images can lead to depth perception errors. This result motivates a careful investigation of differences between types of stereo display hardware.

Finally, some remarks about two parameters that are very important for the setup of stereoscopic display software: the interocular distance and the viewing distance between the observer and the monitor screen. First, the distance between perspective viewpoints used to compute the two views of a stereo image must be equal to the viewer's interpupillary distance. Unfortunately, the distance between human eyes varies between 50 and 70 mm. Hence, the visual impression of a view computed for one observer is not necessarily the same for another. Utsumi, Milgram, Takemura, and Kishino (1994) demonstrated that an interpupillary-distance mismatch results in a certain perceptual error of an object's depth. Therefore, system designers have to make sure that the interocular-distance parameter is an interactively adjustable parameter. It is unfortunate that most systems are not concerned with it. An interpupillary mismatch can also be caused by convergence of the eyes. It is difficult to avoid this type of mismatch, but we should be aware of it because it may cause a large depth estimate error in certain VR environments, for example in an environment using HMDs (Utsumi et al., 1994).

Second, the perceived depth of a point in a stereoscopic display should change with the viewing distance from the monitor screen. When observers move closer to the display, the horizontal parallax increases and, when they move further away, the horizontal parallax decreases for a fixed point. Often stereoscopic display software assumes a fixed distance from the viewing plane. This assumption determines a fixed horizontal disparity for an observed point. Hence, when viewers move perpendicularly to the screen, the perceived shape of the observed object changes. Moving further from the screen produces an elongated and moving closer to the screen produces a contracted image. Hence, there is an optimal position at which the image is perceived at a correct scale.

4.3 User-driven visualization environments

It is the declared goal of a good visualization to gain a better understanding of phenomena that underlie a set of data. Often we find the most expressive and graphically convincing visualization by trial and error. Furthermore, one visualization might provide new insights to one user but not to another. There are many factors influencing the quality of visualizations. Some of these factors, such as the users, and their perceptual experience, data, and their associated perceptually based visualization rules, are the subject of the discussion in this section.

When mapping numerical data to visual features of an image, we need to exercise care. Before engaging in the visualization of data, specialists should study the problem domain they design the visualization for. Only then it is possible to create meaningful visualizations that fit the semantics of the data to be visualized and to minimize the time factor in the development process. As we will show below, it is desirable to furnish the visualization system itself with background information about the user and the data set.

Domik and Gutkauf (1994) suggest a user model that incorporates collective information of a particular user. They assume that every user can be described in terms of past experiences, education, culture, and individual limitations through explicit or implicit modeling of potential users. The former is done by interviewing users or by asking them to fill out questionnaires. For the latter, users are observed in their use of the system performing special tasks. Domik and Gutkauf (1994) studied user performance in tasks related to color perception, mental rotation, and fine motor control. The study of the latter took also the age and gender of the people tested into account. We think that the model could describe other user parameters as well, such as the ability to discriminate shape, the use of different senses, or the ability to perceive depth information based on apparent size (Sacher et al., 1997). Although the results of these ability tests are well understood, it still remains an open question how to integrate these observations into a real visualization system and, therefore, how to get a little step closer to the goal of implementing a visualization system as an extension of our sensory system.

Another component of this discussion of user driven visualization environments is the data model, which serves to organize the information in the data to be visually represented, and its associated visualization rules. Depending on the structure of data, visualization software often supports a variety of procedures for mapping, manipulating, and rendering data. Rogowitz and Treinish (1993) suggest visualization operations based on metadata, that is higher-level representations of the data, such as the dynamic range of data or geometric relationships between objects in an image. Visualization rules, invoked by the user, can then provide guidance on the selection of these operations based on human perception and cognition. These rules constrain the way in which the data would be mapped onto visual dimensions and determine if the visualization represents the structure of the data faithfully and isomorphically or if the visualization emphasizes certain features by transforming the data. For example, an array of scalars, such as altitude data of a surface, could be presented using discrete strategies, e.g. gray-scale or color 2D or 3D scatter plots, or continuous strategies, e.g. gray-scale or color contour lines or region plots. As pointed out earlier, the chosen visualization strategies have to ensure that structures in the data are veridically presented to the user and that perceptual artifacts are not erroneously interpreted by the user as data features. In our example, displaying the continuous variable altitude in a visualization that uses a discrete colormap might introduce visual artifacts. By now, we know that it is important to understand how visualization observers process the visual information presented to them and how variations in this representation can affect its interpretation. The representations are often mathematically identical but look different perceptually to the observer. Hence, different representations of the data can lead to differences in data analysis and interpretation. Tufte (1983) and other researchers emphasize that different representations of the same data can interact perceptually and therefore, complement each other.

Rogowitz and Treinish (1993) suggest that a set of rules should control the mapping from physical to perceptual dimensions. They propose two classes of perceptual rules. The first class encompasses rules that ensure an isomorphic mapping from data to visual features that are perceptually invariant, for example size and color. Hence, this class of rules ensures a faithful data representation in the visualization. A second class of rules helps to highlight features, inherent in the data, to attract the observer's attention for a certain phenomenon in the data by changing the scale or partitioning the image. Using such rules, the user of a visualization system receives help to select visualization operations that map the data onto appropriate visual dimensions with the goal of minimizing perceptual artifacts. Furthermore, the user receives guidance to create the appropriate visualization.

A rule based visualization system must explicitly integrate principles of human vision, perception, and cognition in the visualization process. Rogowitz and Treinish (1993) propose a data model which requires self-describing data structures. These data structures show the logical and physical organization of the data and their relation to each other. In many application builders, for example IBM's Data Explorer or SGI's

Iris Explorer, the data flow to visualization operations is driven by the structure of the data. Adding a complementary layer for providing rule-based guidance would allow us to create rule-based visualizations based on the properties of the metadata, which is either part of the data or may be computed as needed. Therefore, the user does not interact with the data itself but with a higher-level characterization of the data. Furthermore, this interaction is controlled by perceptual rules. Based upon the metadata, a standard operation, which might be already part of the system, can be invoked to create a rule-constrained visualization.

The described approach does not prevent the scenario where the invocation of one rule has an effect on other rules. Sharing metadata across different data structures could resolve this conflict. This requires a sophisticated rule management that, with the interaction of the user, helps to ease or resolve conflicts among the rules. Another task of this management system could be to add new rules including their enabling metadata to the system. In general, the design and management of the perceptual rules should lead to a taxonomy of metadata and an associated class of rules and system operations. It will help us to better characterize and visualize data.

The implementation of rule-based visualizations, as introduced by Rogowitz and Treinish (1993), does not suggest an expert system approach. In general, expert systems are used for well-defined problems in a particular area (Rich and Knight, 1991). Based on batch-like processing, expert systems in visualization use a knowledge base of rules to create the final visualization. In this approach, the user has little control over the visualization process. Furthermore, visualization tasks have often an exploratory nature, which mean incomplete knowledge about the problem domain. In addition, as we pointed out earlier, the process of scientific investigation is an iterative process in which the investigator's knowledge changes and the problem domain is modified. In contrast, the model, which was described above, leaves control in the users' hand and provides interactively guidance to help the user select visualization operations. The idea here is to minimize perceptual artifacts and to create an appropriate visualization.

5 Challenges—the Role of Stereoscopic Visualization

In this chapter, we integrate the knowledge presented in the chapters before. Consider the following stereoscopic visualization. We want to create a visualization tool for geologists that generates 3D models from information in stereo images. This can be done by correlating all of the points on a stereo image and adding a few control points, which could then be used to scale the image. However, until recently this has been a challenging task for a computer to perform on an arbitrary real scene because of insufficient computing power. Given the difficulty of a computer doing this without human intervention, an interactive approach could be advantageous. Once all points are correlated, the parallax and the relative (x,y,z) position of all points on the images (absolute scale and position is known if we have control points) are simply calculated with geometry. One could then use surface rendering to drape the image information onto a 3D model of it and thus "see" and manipulate a 3D scene on the screen. Below, some of the challenges in stereoscopic visualization are discussed using this visualization problem to motivate possible research approaches.

5.1 Some important design aspects of stereoscopic visualization tools

When observing visualizations, it is not enough to perceive the information. In addition, we need also to understand the information presented to us. Our perception and understanding depend on our physiological and perceptual abilities, but also on our age, gender, culture, experience, and on a particular task domain. Are there other factors that need to be examined? Also, what exactly is their influence on the interaction with the planned system? The mentioned factors should make it clear that the creation of visualizations cannot be the effort of a system designer alone, but needs the combined effort of scientists from different fields. This, of course, asks for the active engagement of researchers, system developers, as well as users in the design process. The question that needs to be answered is how to, first, evaluate applicable existing research and experience, and second, integrate this knowledge into the visualization tool design. Past HCI research has shown that people develop deficiencies in perception (e.g. lower visual acuity and reduced contrast sensitivity), and cognition with age. Can we address these diverse needs in our system design?

As pointed out, results from research in perception have direct implication for the visualization process, and it is very important to find a way to incorporate this research into current and future VR visualization systems. We discussed why our knowledge about depth perception should be integrated in the design of stereoscopic displays and stereoscopic display technology. But, to date, the use of this technology is still somewhat limited. We believe that this technology would be more accessible for visualization if we had a better understanding of how we perceive and understand the displayed information. A gain in knowledge will also foster improvements of the visualization process in general.

Indeed, users of visualization software are usually not very progressive when it comes to visualization techniques. Sometimes techniques can be shared across specific areas, but it is often the case that current techniques are not sufficient for the application we have in mind. New ways of viewing data are needed, but the programming costs to develop them are often high. Nevertheless a three-folded approach should be kept in mind: "technology-driven (what we can do), perception-driven (what makes sense), and task-driven (what users want) (Encarnacao et al., 1994, p. 468)". We argue that the development of stereoscopic (and VR) applications is more than the just the utilization of (a new) technology. However, we need a structured, scientific approach to its use rather than a market-driven one.

Users should be in control of visualization tool parameters because they decide about their task performance. How can we give them more control over these parameters? One immediate thought is that the tool defaults should be able to be changed interactively by the user. For instance, we now know that interpupillary distance and the distance to the screen are crucial for how we perceive a stereoscopic image. Rather than hardcoding these two parameters, as unfortunately done in many systems, default values should be provided which easily can be modified and stored. Other factors that are part of our user model could be treated similarly. Furthermore, we could develop a set of benchmark tests to determine the ability of an individual user and to initialize further parameters of our system. Some of the tasks that determine the ability to use the proposed visualization tool for geologists might include the judgment of virtual distance, the estimation of size of virtual objects, and the perception and discrimination of colors.

Many requirements have to be met in visualization design. Research in user interface design has stated design principles in terms of understandabilty and usability. If the design of visualization tools is user centered then we need to provide a good conceptual model of our tool and visible feedback for the guidance of users. Many techniques, such as domain specific user studies, videotaping, interviews, and controlled experiments, have been developed to support the design process of user interfaces. What prevents us from applying these techniques to the development of visualization tools is the time and cost, although some might argue that it is missing commitment of the developer to improving things for the user.

Finally, we want to mention that stereoscopic display and virtual environments in general, offer a superior experimental environment to study perception because we are able to control many parameters that determine our perceptual experience.

5.2 Interacting with three-dimensional visualizations

Visualizing scientific data of three and more dimensions requires new techniques to explore and analyze the data developed. The visualization of phenomena of higher complexity necessitates interactive exploration, and more dimensions of the data mean more degrees of freedom to control making the exploration of higher-dimensional data sets more difficult. Key issues for a meaningful exploration of these data are the choice of visualization techniques, well-designed interaction mechanisms, and, we believe, a good stereoscopic display technology.

Of course, to take full advantage of these new technologies, we are in need of truly multimodal interfaces that integrate real-time video, audio I/O devices, and force-feedback devices. A multi-modal interface can, for certain tasks, lead to enhanced human performance. At a first glance, such a multi-modal interface seems to be unnecessary for our planned visualization tool. Although, there might be tasks for which it is not clear a priori whether voice input, direct manipulation, or gesture input is more appropriate. If we decide for a gesture-based metaphor a better investigation of the used metaphor is desired remembering the problems pointed out earlier.

Graphical user interfaces used for the control of 3D visualization tools are still based on two-dimensional interaction paradigms. In order to manipulate virtual environments more effectively, new interaction metaphors other than the classic 2D window or toolbars with their old affordances and metaphors are needed. What are the affordances that assist the user in our virtual world as much as they would in the real world? Similarly, interaction devices used during the visualization process in virtual environments should be based on an interface other than the 2D mouse. Mouse-based input in our 3D visualization tool, which is for instance necessary for the measurement of distance in the stereo image between two marked points, leaves the problem of mapping the two degrees of freedom of the mouse to the three degrees of freedom of the VR environment unresolved. Many novel 3D-interaction devices exist on today's market. Most of them have not been tested and evaluated regarding their usability in a traditional 3D or stereoscopic environment. We believe that the final design of a new 3D-interaction device that allows an effective control of positions and orientations of objects in three dimensions, is still an open question.

More research has to be conducted to better understand the human factors involved in the design of and interaction with virtual environments. In the case of the geology visualization tool, we need to understand better what makes the user feel comfortable with this system, such as representation of scale, orientation, and time and which help facilities should be provided. In general, the development of new interfaces must be based on a number of different areas, such as

- 3D (interactive) graphics, providing illumination models and depth cuing mechanisms, and animations;
- perceptual psychology, which can answer questions such as how humans process color, depth representations and motion, and
- cognitive science, which is concerned with mental models and problem solving.

As described earlier, previous research showed clear advantages of stereoscopic displays over conventional two-dimensional displays in some tasks. The projection of 3D objects onto 2D screens introduced many perceptual ambiguities. Traditional viewing techniques such as rotation, or different viewing angles were able to reduce these ambiguities. But a combination of these techniques with stereo always produced the best result in disambiguating 3D structures. However, many open questions with regard to perceptual issues, navigation in these virtual worlds, and technical limitations of the current technology need to be addressed by future research. Questions, which the research community often has not started to address in detail, include:

- What is the crucial perceptual information within stereoscopic applications and how task specific is it?
- When and in what tasks should we use a particular viewing paradigm (e.g. egocentric rather than an exocentric viewpoint)?
- How do limitations and artifacts of current stereoscopic display systems affect our performance and perception?

It seems that the main experience associated with stereoscopic display environments is entertainment value, that is the viewing of a 3D image. Although there are studies that examine the benefits of stereoscopic displays in several spatial tasks, little research has been published about the conditions when stereoscopic displays are superior over conventional two-dimensional displays for a real task domain (e.g. geology). Such research has the potential to highlight criteria for display design and for the design of virtual worlds in general.

VR proposes a new interface for human-computer interaction. At its low end, it provides an inexpensive, often head-coupled stereoscopic display environment. At it's high end, we find environments, where the keyboard is replaced by new input devices, such as the data glove, and the computer screen is replaced by head-mounted displays, or projection-based stereoscopic display systems that are powered by expensive supercomputers. Using VR in one or the other form, we are looking for an interface that requires as little attention as possible but allows the user to interact in three dimensions as unambiguous as possible. For many reasons, the design and the implementation of an optimal VR system may not be possible today. Often certain technologies are immature, or if they exist, they might be too expensive.

But despite these problems, VR interfaces are able to create 3D computer-generated environments. They are capable of displaying objects in three dimensions without the known ambiguities of 2D displays and attempt to provide a simple way to select and manipulate these parts of the virtual world. In our example, stereoscopy allows us to explore rapidly and easily a complex data environment in a more natural way using directly controlled visualization tools. Of course, the immaturity of the technology constraints currently the type of applications because the time to render the visualization is often to high to ensure the feeling of immersion and accurate user control. The speed of the visualization hardware, the algorithms used to create the visualization, and the size and complexity of the data sets are some of the limiting factors. Therefore, today's 3D visualization environments, which are based on VR technology, are relatively simple regarding the computational and rendering efforts.

The careful reading and comparison of reports on visualization applications in low-end (Arthur et al., 1993; Pausch, 1991) and high-end virtual reality environments (Cruz-Neira et al., 1993) hints that often the financial efforts to build high-tech virtual visualization environments do not seem to be justified. Many applications offer data views only from a distance and do not utilize an inside-out viewing paradigm. Why? Virtual reality research can be done and many applications can be implemented without expensive hardware. Affordable VR visualization tools, such as (head-coupled) stereoscopic visualization environments, would increase the number of people actively using this technology enabling VR research and applications to mature.

To decide about tasks where VR is clearly superior or uniquely capable, we have to continue the recently begun careful studies and tests of application environments. What are the tasks that might benefit from VR worlds and why are others performed poorly in these environments? For example, the geology visualization system does not seem to benefit greatly from immersion and, therefore, a stereoscopic display is sufficient for this application. The geologists emphasized that they already use stereo images for certain tasks, but are unable to find an affordable computer-based tool that could be easily tailored to their needs. Implementing a real world task, the use of VR has to improve task performance when compared to alternative approaches.

Thus far, VR has fallen short on real-world applications implemented in a large scale. But it promises a bright future for scientific visualization for multi-dimensional data. We need now to deliver working tools, which researchers are willing to use.

5.3 The user is a central part of the visualization process

If we ensure that evolving (stereoscopic) visualization tools should not remain technology-driven but accommodate more the requirements of their users, and if we agree that the process of scientific investigation should be computer-assisted as well as human driven (Watson, 1990) then there is clearly a need for models which describe the main components of the visualization process and the interface between them, including the users and their behavior. These models can clarify system design requirements by stating expectations of and assumptions about the user in general or very specific terms.

Robertson, Earnshaw, Thalmann, Grave, Gallop, and De Jong (1994) specify several criteria to differentiate users and their needs. Using the involvement of the user in the computing process as one criteria, we can distinguish between end-users, who use software as it is without modifying it, and programmers at the interface, system level, or application level.

Another classification scheme takes the different application domains into account. The end-user requirements for visualization tools will most certainly be different for an application in psychology compared with an application in geology. The domain often defines visualization techniques, interface components and more. The different types of visualization tools itself could be used to classify the user's need. Some users might work in a domain where they utilize only 2D visualization techniques. Others are in need of advanced 3D or stereo rendering techniques using different output media. It seems that, at different stages of their work, users will be in need of different visualization techniques and tools. Under what conditions do we need 3D stereoscopic worlds rather than 2D environments?

Looking at how visualization tools are used, we realize that users operate them in different ways. Batch mode postprocessing of visualization results with or without supervision is one operation mode. For our visualization tool, we suggest an interactive, supervised correlation algorithm (i.e. a person could indicate approximate areas of match), so the computer doesn't have to search too far afield, and "help" by interactively "telling" the computer that there is a gap or a shadow or something where the computer gets confused, and then it could proceed.

Users are involved in the visualization process more heavily by observing the visualization result as it evolves or by actively steering simulations or choosing visualization techniques and system parameters. What are user characteristics that might change the experience in virtual worlds? These characteristics can influence user performance tremendously. One of the characteristics is the level of experience of users (novice versus expert paradigm). We hypothesized that this particular user characteristic might even alter how we perceive information in a stereoscopic display (Sacher et al., 1997). Furthermore, the level of experience determines user skills and, hence, performance.

Modeling the user as one core component of the visualization process is a two-folded problem. First, we need to be clear about the perceptional components of the visualization process in relation to a specific user. Steered by our anatomical setup, what we see is often what we want or what we have learned to see. Second, there is a need for the developer of visualization tools to understand the problem domain thoroughly and the perceptual rules invoked when activating a specific visualization technique. The provided visualization technique determines in a way how (correctly or incorrectly) we interpret and understand data sets and their underlying phenomena.

References

Akka, R. (1991a). Writing stereoscopic 3D graphics software using Silicon Graphics GL. . San Rafael: StereoGraphics Corporation.

Akka, R. (1991b). Software concepts for stereoscopic 3D graphics. San Rafael: StereoGraphics Corporation.

Arthur, K. W., Booth, K. S., & Ware, C. (1993). Evaluating 3D Task Performance for Fish Tank Virtual Worlds. *ACM Transactions on Information Systems*, *11*(3), 239-265.

Balakrishnan, R., Baudel, T., Kurtenbach, G., & Fitzmaurice, G. (1997). *The rockin' mouse: integral 3D manipulation on a plane*. Paper presented at the CHI97, Atlanta.

Becker, S., & Patterson, R. (1997). *Temporal differences between crossed and uncrossed stereopsis related to occlusion*. Poster presented at the Association for Research in Vision and Ophthalmology 97, Ft. Lauterdale, FL.

Beverly, K. L., & Regan, D. (1982). Texture changes versus size changes as stimuli for motion in depth. *Vision Research*, *23*(12), 1387-1400.

Braunstein, M. L., Andersen, G.J., Rouse, M.W., Tittle, J. (1986). Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Perception and Psychophysics*, *40*(4), 216-224.

Brodlie, K. W., Carpenter, L.A., Earnshaw,, R.A., Gallop, J.R., Mumford, A.M., Osland, C.D. and Quarendon, P. (Ed.). (1992). *Scientific Visualization - Techniques and Applications* : Springer Verlag.

Brooks, F.P.Jr. (1986). *Walkthrough -- a dynamic graphics system for simulating virtual buildings*. Paper presented at the ACM Workshop on Interactive 3D Graphics.

Brown, M. E., & Gallimore, J. J. (1995). Visualization of Three-Dimensional Structure During Computer-Aided Design. *International Journal of Human-Computer Interaction*, *7*(1), 37-56.

Bryson, S. (1994). Real-time exploratory scientific visualization and virtual reality. In L. Rosenblum, Eearnshaw, R., Encarnacao, J., Hagen, H., Klimenko, S., Nielson, G., Post, F., Thalmann, D. (Ed.), *Scientific visualization: advances and challenges* (pp. 65 - 86). San Diego: Academic Press Inc.

Bryson, S. (1996). Virtual reality in scientific visualization. *Commnications of the ACM*, *39*(5), 63-71.

Buelthoff, H., Fahle, M., Wegmann, M. (1991). Perceived depth scales with disparity gradient. *Perception*, *20*, 145-153.

Burt, P., & B.A., J. (1980). *A disparity gradient limit for binocular fusion. Science*, *208*, 615-617.

Buxton, W. (1983). Lexical and pragmatic considerations of input structures. *Computer Graphics*, *17*(1), 31-37.

Card, S. K., Mackinlay, J. D., & Robertson, G. G. (1990). *The design space of input devices*. Paper presented at the CHI'90.

Card, S. K., Mackinley, J. D., & Robertson, G. G. (1991). A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems*, *9*(2), 99-122.

Chen, M., Mountford, J., & Sellen, A. (1988). *A study in interactive 3D rotation using 2D control devices.* Paper presented at the SIGGRAPH88.

Cruz-Neira, C., Leigh, J., Barnes, C., Cohen, S., Das, S., Englemann, R., Hudson, R., Papka, M., Siegel, L., Vasilakis, C., Sandin, D.J., DeFanti, T.A. (1993). *Scientists in wonderland: a report on visualization applications in the CAVE virtual reality environment.* Paper presented at the IEEE Symposium on Research Frontiers in Virtual Reality.

Davis, E. T., Corso, G. M., Barfield, W., Eggleston, R. G., Ellis, S., Ribarsky, B., & Wickens, C. D. (1994). Human Perception and Performance in 3D Virtual Environments. , *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Vol. 1, pp. 230-234).

Domik, G. O., & Gutkauf, B. (1994). *User Modeling for Adaptive Visualization Systems.* Paper presented at the Visualization '94, Washington.

Earnshaw, R. A., & Wiseman, N. (1992). *An Introductory Guide to Scientific Visualization*: Springer Verlag.

Earnshaw, R. (1994). Fundamental approaches to interactive real-time visualization systems. In L. Rosenblum, Eearnshaw, R., Encarnacao, J., Hagen, H., Klimenko, S., Nielson, G., Post, F., Thalmann, D. (Ed.), *Scientific visualization: advances and challenges* (pp. 55 -65). San Diego: Academic Press, Inc.

Encarnacao, J., Foley, J., Bryson, S., Feiner, S. K., & Gershon, N. (1994). Research issues in perception and user interfaces. In L. Rosenblum, R. A. Earnshaw, J. Encarnacao, H. Hagen, A. Kaufman, S. Klimenko, G. Nielson, F. Post, & D. Thalmann (Eds.), *Scientific visualization-advances and challenges* (pp. 467-472). San Diego: Academic Press, Inc.

Foley, J. D., van Dam, A., Feiner, S.K., Hughes, J.F. (1990). *Computer Graphics.* (second ed.). New York, NY: Addison-Wesley Publishing Company, Inc.

Friedhoff, R. M., & Benzon, W. (1989). *Visualization: the second computer revolution.* New Yourk, NY: Harry N. Abrams, Incorporated.

Gallimore, J. J., & Brown, M. E. (1993). Visualization of 3D Computer-Aided Design Objects. *International Journal of Human-Computer Interaction, 5*(4), 361-382.

Gershon, N. (1994). From perception to visualization. In L. Rosenblum, Eearnshaw, R., Encarnacao, J., Hagen, H., Klimenko, S., Nielson, G., Post, F., Thalmann, D. (Ed.), *Scientific visualization: advances and challenges* (pp. 129 - 142). San Diego: Academic Press Inc.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghtin Mifflin.

Goldstein, E. B. (1989). Perceiving depth and size. In E. B. Goldstein (Ed.), *Sensation and Perception* (pp. 227-273). Belmont, CA: Wadsworth Publishing Company.

Gregory, R. L. (1966). *Eye and brain.* (1^{st}. ed.). New York, NY: McGraw-Hill.

Hamming, R. (1962). *Numerical Methods for Scientists and Engineers.* New York, NY: MCGraw-Hill.

Hinckley, K., Pausch, R., Goble, J. C., & Kassel, N. F. (1994). *A survey of design issues in spatial input.* Paper presented at the UIST'94.

Hinckley, K., Tullio, J., Pausch, R., Proffitt, D., & Kassel, N. (1997). *Usability analysis of 3D rotation techniques.* Paper presented at the UIST'97, Banff, Alberta, Canada.

Hodges, L. F. (1991). Basic principles of stereographic software development. *Stereoscopic Displays and Applications II, SPIE, 1457*, 9-17.

Hodges, L. F., McAllister, D.F. (1990). Rotation algorithm artifacts in stereoscopic images. *Optical Engineering, 29*(8), 973-976.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology, 54*, 21-37.

Ittelson, W. H. (1952). *The Ames demonstrations in perception*. Princeton, NJ: Priceton University Press.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31*(7/8), 1351-1360.

Keller, P. R., & Keller, M. M. (1993). *Visual cues: practical data visualization*. Los Alamitos, CA: IEEE CS Press.

Lipton, L. (1982). *Foundations of the stereoscopic cinema*. New York, N.Y.: Van Nostrand Reinhold Company.

Lipton, L. (1990). Stereo vision on your workstation. *Mechanical Engineering, 112*(3).

Lipton, L. (1991a). The evolution of electronic stereoscopy. *SMPTE journal: publication of the society of motion picture and television engineers, 100*(5), 332-336.

Lipton, L. (1991b). *The CrystalEyes handbook*. San Rafael, CA: StereoGraphics Corporation.

Livingstone, M. S. (1988). Art, illusion and the visual system. Scientific American. *Scientific American, 256*(1), 78-85.

Gershon, N. (1994). From perception to visualization. In L. Rosenblum, Eearnshaw, R., Encarnacao, J., Hagen, H., Klimenko, S., Nielson, G., Post, F., Thalmann, D. (Ed.), *Scientific visualization: advances and challenges* (pp. 129 - 142). San Diego: Academic Press Inc.

Gobbetti, E., Balaguer, J.F., Thalmann, D. (1993). *VB2: A framework for interaction in the synthetic world*. Paper presented at the UIST, Atlanta.

MacAdam, D. L. (1954). Stereoscopic perceptions of size, shape, distance and direction. *SMPTE Journal, 62*, 271-293.

McClain, J. E. (1989). Hue and Disparity Interactions in Advanced Stereoscopic Aircraft Displays. , *Proceedings of the Human Factors Society 33rd Annual Meeting* (Vol. 2, pp. 1422-1426).

McCormick, B., DeFanti, T. A., & Brown, M. D. (1987). Visualization in scientific computing. *ACM SIGGRAPH Computer Graphics, 21*(6).

Mercurio, P. J., & Erickson, T. D. (1990). Interactive Scientific Visualization: An Assessment of a Virtual Reality System. , *Proceedings of IFIP INTERACT'90: Human-Computer Interaction* (pp. 741-745).

Minghim, R., Forrest, R. (1995). *An illustrated analysis of sonification for scientific visualization*. Paper presented at the Visualization '95.

Muehlbach, L., Boecker, M., Prussog, A. (1995). Telepresence in videocommunications: a study on stereoscopy and individual eye contact. *Human Factors: the journal of the Human Factors Society, 37*(2), pp. 290.

Ramachandran, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours. *Pereception & Psychophysics, 39*(5), 361-373.

Rich, E., & Knight, K. (1991). *Artificial Intelligence.* (second ed.). New York: McGraw-Hill, Inc.

Owens, D. A., & Libowitz, H. W. (1975). Chromostereopsis with small pupils. *Journal of the Optical Society of America, 65*, 358-359.

Patterson, R., Short, G. L., & Moe, L. (1989). Stereoscopic Depth Sensitivity Differences between the Crossed and Uncrossed Directions. , *Proceedings of the Human Factors Society 33rd Annual Meeting* (Vol. 2, pp. 1427-1429).

Patterson, R., Moe, L., Hewitt, T. (1992). Factors that affect depth perception in stereoscopic displays. *Human Factors: the journal of the Human Factors Society, 34*(6), 655-668.

Pausch, R. (1991). Virtual Reality on Five Dollars a Day. , *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems* (pp. 265-270).

Predebon, J. (1993). The familiar size-cue to distance and stereoscopic depth perception. *Perception, 22*(8), 985-996.

Robertson, P. K., Earnshaw, R. A., Thalmann, D., Grave, M., Gallop, J., & De Jong, E. M. (1994). Research issues in the foundations of visualization. In L. Rosenblum, Eearnshaw, R., Encarnacao, J., Hagen, H., Klimenko, S., Nielson, G., Post, F., Thalmann, D. (Ed.), *Scientific visualization: advances and challenges* (pp. 479-488). San Diego, CA: Academic Press Inc.

Rogowitz, B. E., & Treinish, L. A. (1993). *An Architecture for Rule-Based Visualization.* Paper presented at the Visualization '93, San Jose.

Sacher, G., Hayes, A., Thornton, I. M., & Sereno, M. (1997). *Apparent size drives depth reversals in stereoscopic displays.* Poster presented at the Association for Research in Vision and Ophthalmology 97, Ft. Lauterdale, FL.

Sollenberger, R. L., & Milgram, P. (1991). A Comparative Study of Rotational and Stereoscopic Computer Graphic Depth Cues. , *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1452-1456).

Southard, D. A. (1992). Transformations for stereoscopic visual simulation. *Computer & Graphics, 16*(4), 401-410.

Treisman, A. (1987). Properties, parts, and objects. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* . New York: Wiley.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Utsumi, A., Milgram, P., Takemura, H., & Kishino, F. (1994). Investigation of Errors in Perception of Stereoscopically Presented Virtual Object Locations in Real Display Space. , *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Vol. 1, pp. 250-254).

Valyus, N. A. (1966). *Stereoscopy.* London, GB: Focal Press.

Venola, D. (1993). *Facile 3D direct manipulation.* Paper presented at the ACM INTERCHI'93 Conference on Human Factors in Computing Systems.

Vince, J. (1995). *Virtual Reality Systems.* New York, NY: Addison-Wesley Publishing Company, Inc.

Vos, J. J. (1960). Some new aspects of color stereoscopy. *Journal of the Optical Society of America, 50,* 785-790.

Wann, J. P., Rushton, S., Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Human Factors: the journal of the Human Factors Society, 35*(19).

Ware, C., Arthur, K., & Booth, K. S. (1993). Fish Tank Virtual Reality. , *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems* (pp. 37-42).

Ware, C., & Osborne, S. Exploration and virtual camera control in virtual three-dimensional environments. *Computer Graphics, 24*(2), 175-183.

Watson, D. (1990). *The state of the art of visualization.* Paper presented at the SuperComputing Europe Fall Meeting, Aachen.

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosphical Transactions of the Royal Society of London*, 371-394.

Wickens, C. D., Mervin, D.F., Lin, E. (1994). Implication of graphics enhancements for the visualization of scientific data: dimensional integrality, stereopsis, motion, and mesh. *Human Factors: the journal of the Human Factors Society, 36*(1).

Yey, Y.-Y., & Silverstein, L. D. (1990). Limits of Fusion and Depth Judgment in Stereoscopic Color Displays. *Human Factors, 32*(1), 45 -60.

Zenyuh, J. P., Reising, J. M., Walchli, S., & Biers, D. (1988). A Comparison of a Stereographic 3D Display versus a 2D Display Using an Advanced Air-to-Air Format. , *Proceedings of the Human Factors Society 32nd Annual Meeting* (Vol. 1, pp. 53-57).