2. VISUAL INTERFACES

A VE imposes a number of requirements for visual displays. The most significant of these are stereoscopic vision and the ability to track head movements and continually update the visual display to reflect the user's movement through the environment. In addition, the user should be surrounded by visual stimuli of adequate resolution, in full color, with adequate brightness, and high-quality motion representations. These requirements are extremely demanding, given the capabilities of current displays and computing platforms, although progress is rapidly being made.

Another major challenge is the provision of display hardware that is not only capable of providing the necessary quality at an acceptable cost, but that minimizes the impact on the user. Currently available displays for immersive VEs typically require the user to wear a head-mounted display (HMD) or some form of special glasses. These introduce a range of new issues, such as ergonomics and health concerns, which are particularly critical in the case of HMDs.

There are a very large number of different techniques for providing stereoscopic vision. These fall into the general categories of HMDs, active (shutter) glasses, passive glasses, and autostereoscopic displays. Present-day HMDs use a technique in which each eye is provided with a separate display, together with optics that magnify the image and allow the user to focus at some depth other than the surface of the display screens. The displays and associated optics are mounted in a helmet type device, often with a position tracker and headphones attached. A new version of an HMD, now in the research stages, is based on the retinal display, in which images are not displayed on a screen, but are created by directing a beam of light (such as from a laser) to the retina of the eye.

A similar, but less encumbering alternative is the use of special glasses. In active glasses, electronic shutters are mounted in the place of the lenses of eyeglasses and, hence, these devices are often called shutter glasses. The shutters are monochrome LCDs that are used to display an opaque image to one eye and a transparent image to the other, continually switching between eyes. The user looks at a cathode ray tube (CRT) monitor or projection screen that shows left and right images as sequential fields, and that also generates a synchronization signal (such as from an infrared emitter) that controls the timing of the shutters. Passive glasses, on the other hand, use an approach in which perspective views for each eye are encoded in the form of either color (for example, red for one eye and green for the other) or polarization of the light, with the "lens" for each eye containing a filter that passes only the appropriate image intended for each eye.

Autostereoscopic displays do not require the user to wear any form of display or special glasses, although a head tracker may be needed. A variety of techniques are used. In some systems, lenses behind or in front of a display screen focus the image so that each eye necessarily sees a different image. In other systems, barriers such as vertical bar in front of the display prevent both eyes from seeing the same image. Another approach uses beams of light to scan a 3-D volume that serves as a projection screen, with the beams reflected to display a pixel at a given coordinate.

The relative strengths and weaknesses of all the different types of devices, and the accompanying technologies, are summarized in Table 1.

Relatively little is known about the conditions that provide a sense of immersion in a virtual environment. While stereoscopic vision is generally considered to be necessary for true 3-D vision and a sense of immersion, the capability of changing the visual image in response to head movement (as occurs in a real environment) may be more critical for 3-D vision than stereopsis in some conditions. For example, in one experiment (Ware, Arthur, and Booth, 1993), subjects made more errors in a task requiring 3-D vision when a stereoscopic display was used without head-coupling than when a head-coupled monoscopic display was used. (The fewest errors of all were made with a head-coupled stereoscopic display.)

This section continues with an introduction to the human visual system that presents those aspects of the visual sense that have an impact on display requirements. Commercially available display devices, intended for use with VEs, are then described, followed by descriptions of on-going R&D in this field. The section finishes with a summary of the current status of technology in VE visual displays and gives some projections for expected advances in the next few years.

2.1 The Human Visual System

The human visual system is very complex and only partially understood. It is clearly powerful with a very high bandwidth and remarkable ability to resolve detail, color, texture, and depth. The visual system also involves substantial capabilities for processing information and complex networks of neurons in both the eye and brain are devoted to visual processing (Hubel, 1963). Vision is generally considered the most dominant sense, and there is evidence that human cognition is oriented around vision, with people often using visual imagery as mediating representations for thought (Kosslyn, 1980; 1994). Thus, it is natural for high-quality visual representations to be considered critical for VEs.

The visual system consists of the eyes, certain pathways and intermediate processing centers that carry visual information from the eyes to the brain, and the visual cortex of the brain. Light enters the eye through the cornea, a transparent bulge, and some proportion of the incoming light passes through the pupil, a circular opening that is similar in form and

Table 1. Visual Display Device Types and Technologies

Technology	Description	Strengths	Weakness
HMDs	Use dual monitors (CRT or LCD) and special optics to present a different image to each eye. The monitors and lenses are mounted in a helmet-type device, usually with a position tracker and ear phones attached.	- Encompassing visual volume - Relative freedom of movement - (CRT-based): Small, high-resolution, high luminance, monochrome displays - (LCD-based): Color displays with low voltages near user's head	- Weight and inertial burden - Single viewer - Bulky optics that introduce distortions - (CRT-based): High voltages near user's head, mechanical or electrical color filtering techniques needed for color, sources of distortion - (LCD-based): Low resolution, slow switching time
Active Glasses	Special glasses with electronic shutters that display images to each eye alternately. Require a special monitor or projection screen that presents left and right eye images as sequential fields.	- Low weight - Multiple viewers possible (static stereo scenes for all but head-tracked user)	- Interocular crosstalk - Not suited for encompassing visual volume - Effectively halves frame rate
Passive Glasses	Special glasses with filters that pass only the image intended for each eye. Require a special monitor or screen in which perspective views for each eye are encoded in form of color or light polarization.	- Low weight - Multiple viewers possible (static stereo scenes for all but head-tracked user) - Inexpensive glasses (though projection display may be expensive)	- Not suited for encompassing visual volume - Poor contrast - Color coding results in eye fatigue
BOOM	Use dual monitors (CRT or LCD) and special optics to present a different image to each eye. The monitors and lenses are mounted in device suspended from a boom in front of the user.	- Weight counter-balanced - Low latency - Ease of switch to keyboard operation	- User movement restricted by mechanical linkages
Autostereoscopic Displays	Common approaches use either lenses positioned behind or before a display screen, or physical barriers in front of the display to cause each eye to see a different image.	- Unencumbering - Multiple viewers possible - Ease of switch to keyboard operation	- Low resolution - Slow switching time - User movement restricted to limited viewing area - Not suited for encompassing visual volume

function to the aperture of a camera. Muscles in the middle of the iris (the colored part of the eye) contract to increase or decrease the size of the pupil. Light that passes through the pupil enters the crystalline lens, a transparent structure that has muscles surrounding it that can rapidly alter its shape, allowing the eye to focus on particular objects, a process known as accommodation. Images are refracted by the lens and projected onto the retina, a thin layer of neural tissue that makes up most of the eye's interior (Davson, 1989).

The retina is often thought of as analogous to the sensor in a television camera that converts light to electricity and does perform this function, but it is also a very complex visual information processing system whose function goes far beyond creating electrical impulses. The structure of the retina is complex, and consists of several layers, with one layer devoted to photoreceptors, and others to concentrating and processing the output of the photoreceptors. The photoreceptor layer itself is relatively complex: there are two main types of photoreceptors, rods, of which there are approximately 120 million, and cones, of which there are approximately 8 million. Rods are used primarily for night vision, have poor sensitivity to detail, and are not sensitive to color, though they are extremely sensitive to low levels of light. Cones have good resolution for detail and are sensitive to color. In fact, there are three different kinds of cones, known as blue cones (with a peak sensitivity at 435 nm), red cones (peak at 565 nm), and green cones (peak at 535 nm). The rods and cones are not at all equally distributed in the retina: the fovea, an area of the retina upon which the central image is projected, has a very heavy concentration of cones and very few rods, while the periphery, or remainder of the field, has a heavy concentration of rods but few cones, with the density of cones decreasing with the distance away from the center of the fovea. As will be discussed later, it is possible to make use of this in the design of visual displays that economically display information in color with high resolution in the fovea and in black and white at low resolution in the periphery.

Complex circuitry in the retina, the lateral geniculate nucleus (a structure between the eye and the brain that does preprocessing), and the visual cortex of the brain perform a variety of processing. Some of this is concerned with color, while other circuitry is concerned with shape. In particular, there are layers of neural tissue that process information so as to identify increasingly abstract information. Thus, lower-level layers detect edges (with some neurons sensitive to horizontal edges, for example, and others vertical), and higher level layers detect more abstract shapes, such as curves that make up objects. Human vision also is highly sensitive to both depth and motion perception. The visual system uses a complex variety of information to determine the depth of an object, such as binocular disparity and linear perspective cues.

The field of view is the angle that an eye, or pair of eyes, can see in either the horizontal or vertical dimension. The total horizontal field of vision of both human eyes is about 180° without eye movement or, allowing for eye movements to the left or right, the total field of vision possible without moving the head is 270°. The vertical field of vision is typically over 120°. While the total field is not necessary for a user to feel immersed in a visual environment, there is a belief among some in the community that at least 90°, and perhaps 110°, is necessary for the horizontal field of vision.

Visual acuity is the ability of the eye to resolve two stimuli separated in space. This measure is significant in that it has implications for image resolution: it is desirable for resolution to be sufficiently high that the ability of the eye to resolve stimuli, rather than the resolution of an image being displayed, is the limiting factor. Visual acuity depends significantly on both luminance levels and whether the stimuli is presented in the fovea or the periphery, with a difference of more than 20:1 between the high acuity seen with bright light in the fovea and the poor acuity resulting from dimly lit stimuli presented in the periphery (Mandelbaum and Sloan, 1947). In general, this reflects the much greater visual acuity for cone cells as opposed to rods. Assessments of visual acuity vary substantially depending upon whether they are determined by calculating the size of the retinal receptors or experimentally by psychophysical measurements, with results ranging from 0.5 to 20 seconds of arc (Davson, 1989). It is common, however, in discussions of visual acuity and its implications for display resolution, to use the more conservative figure of 30 seconds of arc for the smallest resolution visible (McKenna and Zeltzer, 1992).

Visual simulations that work by rapid successive presentations of images to the eye—as in the case of motion pictures, television, or computer-controlled displays—should preferably have successive frames presented at or above a certain rate. This rate is the critical fusion frequency, the point at which stimuli are perceived as a continuous stimulation (as fused) rather than distinct successive images (Davson, 1989). In general, the greater the luminous intensity of a stimuli, the higher the frequency at which successive images must be presented to avoid flicker. In the fovea, the critical fusion frequency is generally proportional to the logarithm of the luminance of the stimuli over a wide range (0.5 to 10,000 trolands). At high luminances, the critical fusion frequency is about 50-60 Hz, while at very low luminances it may be as low as 5 Hz. In addition, the critical fusion frequency is proportional to the size of the area of the retina in which the image falls, as well as other factors (Landis, 1954). While flicker is undesirable—it is annoying, makes perception more difficult and presumably disturbs the sensation of immersion—there is typically a trade-off needed between image complexity and susceptibility to flicker in systems with fixed computational power, and in some applications it may be preferable to tolerate flicker at least some of the time to gain increased scene complexity. Under most conditions, a 60 Hz refresh frequency (used for television in the United States), will result in an absence of flicker. According to McKenna and Zeltzer (1992), a rule of thumb in the computer graphics industry suggests that below about 10-15 Hz, objects will not appear to be in continuous motion, resulting in distraction.

The human eye is sensitive to an extremely wide range of light levels, about 12 logarithmic units. About 6 of these levels are under rod vision, while the other 6 are under cone vision. However, the eye cannot operate at any given time across this entire range: instead, the eye adapts to a given level of light, largely by mechanisms involving the light-sensitive chemicals in the receptor neurons in the retina. Such adaptation is very rapid when light levels increase, but take on the order of minutes or tens of minutes when light levels decrease. For a certain state of adaptation, the eye is sensitive to about two orders of magnitude of brightness.